THE ANNALS *of* MATHEMATICAL STATISTICS

THE OFFICIAL JOURNAL OF THE INSTITUTE OF MATHEMATICAL STATISTICS

Contents

Vol. IX, No. 4 — December, 1938

Manuscripts for publication in the ANNALS OF MATHEMATICAL STATISTICS should be sent to S. S. Wilks, Fine Hall, Princeton, New Jersey. Manuscripts should be typewritten double-spaced with wide margins, and the original copy should be submitted. Footnotes should be reduced to a minimum and whenever possible replaced by a bibliography at the end of the paper; formulae in footnotes should be avoided. Figures, charts, and diagrams should be drawn on plain white paper or tracing cloth in black India ink twice the size they are to be printed. Authors are requested to keep in mind typographical difficulties of complicated mathematical formulae.

Authors will ordinarily receive only galley proofs. Fifty reprints without covers will be furnished free. Additional reprints and covers furnished at cost.

The subscription price for the ANNALS is $4.00 per year. Single copies $1.25. Back numbers are available at the following rates:

Vols. I–IV $5.00 each. Single numbers $1.50.
Vols. V to date $4.00 each. Single numbers $1.25.

Subscriptions, renewals, orders for back numbers and other business communications should be sent to A. T. Craig, University of Iowa, Iowa City, Iowa.

The ANNALS OF MATHEMATICAL STATISTICS is published quarterly by the Institute of Mathematical Statistics.

COMPOSED AND PRINTED AT THE
WAVERLY PRESS, INC.
BALTIMORE, MD., U. S. A.

cs
ts
by
er
t-
on
to
es

ut
st.

25.

m-
va.

he

NOTES ON HOTELLING'S GENERALIZED T

BY P. L. HSU

1. Frequency Distribution When the Hypothesis Tested is Not True

a. THE PROBLEM. Let the simultaneous elementary probability law of the $k(f+1)$ variables z_i and z'_{ir} $(i = 1, 2, \cdots, k; r = 1, 2, \cdots, f)$ be

$$(1) \quad p(z, z') = (\sqrt{2\pi})^{-k(f+1)} | C |^{\frac{1}{2}(f+1)} \exp\left[-\frac{1}{2} \sum_{i,j=1}^{k} c_{ij} \{ (z_i - \zeta_i)(z_j - \zeta_j) + v'_{ij} \} \right],$$

where

$$v'_{ij} = \sum_{r=1}^{f} z'_{ir} z'_{jr} \qquad (i, j = 1, 2, \cdots, k)$$

C stands for the matrix $\| c_{ij} \|$ and $| C |$, the corresponding determinant. It is required to find the elementary probability law of the statistic

$$T = | V' |^{-1} \sum_{i,j=1}^{k} V'_{ij} z_i z_j,$$

where $| V' | = | v'_{ij} |$ and V'_{ij} denotes the cofactor of the element v'_{ij} in the matrix $\| v'_{ij} \|$.

The quantity fT is a generalization of "Student's" t considered by Hotelling [1]*. It is an appropriate criterion to test the hypothesis, say H_0, that the ζ_i in the parent population as given by (1) all vanish. The distribution of T when the hypothesis H_0 is true has already been obtained by Hotelling. But our knowledge of the test is hardly complete unless we know also the distribution of T when the ζ_i do not all vanish. Indeed, only such a knowledge can enable us to control the risk of error of the second kind, i.e. of failure to detect that H_0 is untrue [3, 4].

b. THE SOLUTION. Let H be a $k \times k$ non-singular matrix such that $H'CH = I$, the unit matrix, where H' denotes the transposed matrix of H. Let the sets of variables $(z_1, z_2, \cdots, z_k)$ and $(z'_{1r}, z'_{2r}, \cdots, z'_{kr})(r = 1, 2, \cdots f)$ be subject to the same collineation by means of H, so that

$$\| z_1, z_2, \cdots, z_k \| = \| t_1, t_2, \cdots, t_k \| \cdot H'$$

$$\| z'_{1r}, z'_{2r}, \cdots, z'_{kr} \| = \| t'_{1r}, t'_{2r}, \cdots, t'_{kr} \| \cdot H' \quad (r = 1, 2, \cdots, f)$$

where the t_i and t'_{ir} are the new variables. Let further the quantities τ_i be defined by

---

* References are given at the end of the paper.

$$\| \zeta_1, \zeta_2, \cdots, \zeta_k \| = \| \tau_1, \tau_2, \cdots, \tau_k \| \cdot H'. \tag{2}$$

Then, as is easy to verify, the simultaneous distribution of the new variables will be given by

$$p_1(t, t') = (\sqrt{2\pi})^{-k(f+1)} \exp\left[-\frac{1}{2} \sum_{i=1}^{k} \{ (t_i - \tau_i)^2 + u'_{ii} \} \right], \tag{3}$$

while the statistic T, as a function of the t's, retains the original form:

$$T = | U |^{-1} \sum_{i,j=1}^{k} U'_{ij} t_i t_j \tag{4}$$

where

$$u'_{ij} = \sum_{r=1}^{f} t'_{ir} t'_{jr} \qquad (i, j = 1, 2, \cdots, k),$$

$| U' | = | u'_{ij} |$, and U'_{ij} is the cofactor of the element u'_{ij} in the matrix $\| u'_{ij} \|$. By virtue of (2) we have the following relation between the old and new parametric constants:

$$\sum_{i=1}^{k} \tau_i^2 = \sum_{i,j=1}^{k} c_{ij} \zeta_i \zeta_j . \tag{5}$$

Our problem is thus reduced to finding the derived distribution of T defined by (4) from the parent population given by (3).

We solve this problem by obtaining an expression for the Laplace integral $E(e^{-\beta T})$, i.e. the mathematical expectation of $e^{-\beta T}$ for real non-negative β. A few words are perhaps needed to explain the fact that the Laplace transform of an elementary probability law determines the latter uniquely except on a null set of points. If $f(x)$ is an elementary probability law which vanishes for all negative x and if

$$g(\beta) = \int_0^\infty e^{-\beta x} f(x)\, dx \quad \text{for} \quad \beta \geq 0,$$

then, letting c be any fixed positive constant, we have

$$g(c - \beta) = \int_0^\infty e^{\beta x} e^{-cx} f(x)\, dx$$

for all $\beta \leq c$. We get therefore

$$m_h = \int_0^\infty x^h e^{-cx} f(x)\, dx = \frac{d^h}{d\beta^h} g(c - \beta) \Big|_{\beta=0}, \quad (h = 0, 1, 2, \cdots)$$

the definite integral being obviously finite for all $h \geq 0$. Now a sufficient condition that the set of numbers m_h determines the function $e^{-cx}f(x)$ uniquely, with the exception of a null set at most, is that the latter multiplied by $e^{k\sqrt{x}}$ be summable $(0, \alpha)$ for some positive k (cf. [6], p. 320). Since this condition is trivially satisfied by the function $e^{-cx}f(x)$, this function, and therefore $f(x)$ itself,

must be uniquely determined by the m_h. In other words, $f(x)$ is uniquely determined by its Laplace transform $g(\beta)$. We now proceed to find the Laplace integral $E(e^{-\beta T})$.

Introduce the function

$$g(t, t', \theta, \alpha) = (\sqrt{2\pi})^k \,|\, U' \,|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\left\{\sum_{i,j=1}^{k} u'_{ij}\theta_i\theta_j + 2i\alpha\sum_{i=1}^{k} t_i\theta_i\right\}\right]$$

and write

$$F(t, t', \theta, \alpha) = p_1(t, t')g(t, t', \theta, \alpha),$$

where all the arguments take real values only. For any functions $\varphi(\theta)$ and $\psi(t, t')$ let us write

$$\int \varphi(\theta)\, d\theta = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(\theta)\, d\theta_1 \cdots d\theta_k$$

$$\int \psi(t, t')\, d(t, t') = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \psi(t, t')\, dt_1 \cdots dt_k\, dt'_{11} \cdots dt'_{kf}.$$

We have

$$\int d(t, t') \int |\, F(t, t', \theta, \alpha) \,|\, d\theta = \int p_1(t, t')\, d(t, t') \int g(t, t', \theta, 0)\, d\theta = 1$$

whence we know that

$$(6) \qquad \int d(t, t') \int F\, d\theta = \int d\theta \int F\, d(t, t')$$

On the right-hand side of (6) we find

$$\int p_1(t, t')\, d(t, t') \int g(t, t', \theta, \alpha)\, d\theta = \int e^{-\frac{1}{2}\alpha^2 T} p_1(t, t')\, d(t, t') = E(e^{-\frac{1}{2}\alpha^2 T})$$

while for the integral on the right-hand side of (6) we have

$$\int F\, d(t, t')$$

$$(7) \qquad = (\sqrt{2\pi})^{-k(f+2)} \exp\left(-\frac{1}{2}\sum_{i=1}^{k} \tau_i^2\right) \int \exp\left[-\frac{1}{2}\sum_{i=1}^{k} \{t_i^2 + 2(i\alpha\theta_i - \tau_i\}\right] dt$$

$$\times \int |\, U' \,|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} (\theta_i\theta_j + \delta_{ij})\, u'_{ij}\right] dt',$$

where we mean by the δ_{ij} the quantities

$$\left.\begin{array}{ll} \delta_{ij} = 0 & \text{for } i \neq j \\ \delta_{ii} = 1 & \end{array}\right\} (i, j = 1, 2, \cdots, k)$$

In the equation (7) the integral with respect to the t_i is immediately written down as

$$(\sqrt{2\pi})^k \exp\left[\frac{1}{2}\sum_{i=1}^{k}(\tau_i - i\alpha\theta_i)^2\right]$$

As to the integral with respect to the t'_{ir}, we may evaluate it by the method by which Wilks [7] evaluated the moments of the generalized variance. The result is

$$2^{\frac{1}{2}k}(\sqrt{2\pi})^{kf}\,|\,\theta_i\theta_j + \delta_{ij}\,|^{-\frac{1}{2}(f+1)}\,\frac{\Gamma\frac{1}{2}(f+1)}{\Gamma(\frac{1}{2}(f+1-k))}$$

Making the substitution into (7) we get, after necessary reductions,

$$\int F\,d(t, t') = \frac{(\sqrt{\pi})^{-k}\Gamma(\frac{1}{2}(f+1))}{\Gamma(\frac{1}{2}(f+1-k))}\,|\,\theta_i\theta_j + \delta_{ij}\,|^{-\frac{1}{2}(f+1)} \times \exp\left[-\sum_{i=1}^{k}\{\tfrac{1}{2}\alpha^2\theta_i^2 + i\alpha\tau_i\theta_i\}\right]$$

whence, noticing that $|\,\theta_i\theta_j + \delta_{ij}\,| = 1 + \sum_{i=1}^{k}\theta_i^2$

$$E(e^{-\frac{1}{2}\alpha^2 T}) = \frac{(\sqrt{\pi})^{-k}\Gamma(\frac{1}{2}(f+1))}{\Gamma(\frac{1}{2}(f+1-k))}\int\left(1 + \sum_{i=1}^{k}\theta_i^2\right)^{-\frac{1}{2}(f+1)} \exp\left[-\sum_{i=1}^{k}\{\tfrac{1}{2}\alpha^2\theta_i^2 + i\alpha\tau_i\theta_i\}\right]d\theta \tag{8}$$

Equation (8) gives the Laplace transform of the elementary probability law, $p(T)$, of T. There is no essential difficulty in getting $p(T)$ by inversion directly from (8). Nevertheless, it may be of interest to get $p(T)$ indirectly by identifying the right-hand side of (8) with the Laplace transform of another elementary probability law which is otherwise *known*. For this purpose consider the simultaneous elementary probability law

$$p(x, y) = (\sqrt{2\pi})^{-(f_1+f_2)}\exp\left[-\frac{1}{2}\sum_{i=1}^{f_1}(x_i - \xi_i)^2 - \frac{1}{2}\sum_{j=1}^{f_2}y_j^2\right]$$

and let us find the derived distribution of the statistic

$$L = \sum_{i=1}^{f_1}x_i^2 \Big/ \sum_{j=1}^{f_2}y_j^2$$

As before, we introduce the function

$$g(x, y, \theta, \alpha) = (\sqrt{2\pi})^{-f_1}\left(\sum_{j=1}^{f_2}y_j^2\right)^{\frac{1}{2}f_1}\exp\left[-\frac{1}{2}\left(\sum_{j=1}^{f_2}y_j^2\sum_{i=1}^{f_1}\theta_i^2 + 2i\alpha\sum_{i=1}^{f_1}x_i\theta_i\right)\right]$$

write

$$F(x, y, \theta, \alpha) = p(x, y)g(x, y, \theta, \alpha)$$

and ascertain that

$$\int d(x, y) \int F\,d\theta = \int d\theta \int F\,d(x, y) \tag{9}$$

On the left-hand side of (9) we find

$$\int e^{-\frac{1}{2}\alpha^2 L} p(x, y)\,d(x, y) = E(e^{-\frac{1}{2}\alpha^2 L})$$

while for the integral on the right-hand side of (9), we have

$$\begin{aligned}\int F\,d(x, y) &= (\sqrt{2\pi})^{-(2f_1+f_2)} \exp\left(-\frac{1}{2}\sum_{i=1}^{f_1} \xi_i^2\right) \\ &\quad \times \int \exp\left[-\frac{1}{2}\sum_{i=1}^{f_1} \{x_i^2 + 2(i\alpha\theta_i - \xi_i)x_i\}\right] dx \\ &\quad \times \int \left(\sum_{j=1}^{f_2} y_j^2\right)^{\frac{1}{2}f_1} \exp\left[-\frac{1}{2}\left(1 + \sum_{i=1}^{f_1} \theta_i^2\right)\sum_{j=1}^{f_2} y_j^2\right] dy \\ &= \frac{(\sqrt{\pi})^{-f_1}\Gamma(\frac{1}{2}(f_1+f_2))}{\Gamma(\frac{1}{2}f_2)} \left(1 + \sum_{i=1}^{f_1} \theta_i^2\right)^{-\frac{1}{2}(f_1+f_2)} \\ &\quad \exp\left[-\frac{1}{2}\sum_{i=1}^{f_1} (\alpha^2\theta_i^2 + 2i\alpha\xi_i\theta_i)\right]\end{aligned}$$

Writing

$$f_1 = k, \qquad f_2 = f + 1 - k \tag{10}$$

we get finally

$$\begin{aligned}E(e^{-\frac{1}{2}\alpha^2 L}) &= \frac{(\sqrt{\pi})^{-k}\Gamma(\frac{1}{2}(f+1))}{\Gamma(\frac{1}{2}(f+1-k))} \int \left(1 + \sum_{i=1}^{k} \theta_i^2\right)^{-\frac{1}{2}(f+1)} \\ &\quad \exp\left[-\frac{1}{2}\sum_{i=1}^{k} (\alpha^2\theta_i^2 + 2i\alpha\xi_i\theta_i)\right]\end{aligned} \tag{11}$$

From the identity of (8) and (11) we conclude that T is distributed exactly the same as L with the appropriate "degrees of freedom" f_1 and f_2 given by (10). But the elementary probability law of L has already been derived by P. C. Tang [5]. Using his result we immediately write down the elementary probability law of T:

$$p(T) = e^{-\lambda} \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} \frac{1}{B(\frac{1}{2}f_1 + h, \frac{1}{2}f_2)} T^{\frac{1}{2}f_1+h-1}(1 + T)^{-\frac{1}{2}(f_1+f_2)-h} \tag{12}$$

where f_1 and f_2 are given by (10) and

$$\lambda = \frac{1}{2}\sum_{i=1}^{k} \tau_i^2 = \frac{1}{2}\sum_{i,j=1}^{k} c_{ij}\zeta_i\zeta_j \tag{13}$$

in accordance with (5). The tables of probability integrals prepared by Tang can, of course, be used to suit our purpose.

2. **An Optimum Property of the *T*-Test.** To any reader familiar with the Neyman-Pearson theory of testing statistical hypotheses [3, 4], the theorem stated below may be of considerable interest.

Denote by W the $k(f + 1)$-dimensional space of the z_i and z'_{ir} and let w be any region in W which may possibly serve as a critical region for the rejection of the hypothesis H_0. Let us speak of a critical region w as belonging to the class D if w satisfies the following condition:

$$\int_w p(z, z')d(z, z') = \epsilon + \frac{\alpha}{2} \sum_{i,j=1}^{k} c_{ij}\zeta_i\zeta_j + R \tag{14}$$

where $\epsilon < 1$ is a positive constant independent of the ζ_i, c_{ij} and the region w, α is a constant depending on w only, but not on the ζ_i or c_{ij}, and where R for any given set of values of the c_{ij} is an infinitesimal of at least the third order as all the ζ_i tend to zero.

THEOREM. *Of all the regions belonging to the class D, the particular region which gives the largest possible value to the coefficient α in the equation* (14) *is the region defined by $T \geq T_\epsilon$, where T_ϵ is a constant so determined that the probability, when all ξ_i vanish, of the observed T being not less than T_ϵ is exactly ϵ.*

The significance of the theorem is clear. Every critical region belonging to the class D serves as an unbiased exact test of the hypothesis H_0, ϵ being the preassigned chance of rejecting H_0 if it is true. Further, as is seen from (14), as the ζ_i start to depart from zero, the increased chance of rejecting H_0 due to its falsehood is approximately proportional to the quantity $\Sigma c_{ij}\zeta_i\zeta_j$. The coefficient α therefore measures the power of the critical region w to detect the falsehood of H_0, at least when the departure of the ζ_i from zero is small. Our theorem asserts that in this particular sense the T-test is the most powerful of its kind.

The method of proof is very much the same as that by which Neyman and Pearson proved some of their general theorems concerning unbiased tests. However, as the present theorem has not yet been contained in their more general results, we shall give it a full proof without referring, save in one occasion, to these authors.

PROOF. Write

$$v'_{ij} + z_i z_j = s_{ij} \qquad (i, j = 1, 2, \cdots, k)$$

$$p_0(z, z') = (\sqrt{2\pi})^{-k(f+1)} C^{\frac{1}{2}(f+1)} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}s_{ij}\right] \tag{15}$$

and denote by $p_0(s)$ the simultaneous elementary probability law of the variables s_{ij} derived from (15). Let W_1 be the domain of all possible positions of the point $(s_{11}, s_{12}, \cdots, s_{kk})$ in the $\frac{1}{2}k(k + 1)$-dimensional space.

We know, although we omit the proof of it, that there is no elementary probability law of the variables s_{ij} other than $p_0(s)$ which has the same moments of all orders as those derived from $p_0(z, z')$. It then follows that if $g(s)$ be any summable function of the s_{ij} and if

$$\int_{W_1} \left(\prod_{i,j=1}^{k} s_{ij}^{r_{ij}} \right) g(s) p_0(s) \, ds = 0 \tag{16}$$

for all positive integers r_{ij} or zero, then we must have $g(s) \equiv 0$ except perhaps on a null set of points.

It follows therefore that the identity

$$\int_{W_1} g(s) p_0(s) \, ds \equiv 0 \tag{17}$$

implies the identity $g(s) \equiv 0$ provided $g(s)$ does not involve the parameters c_{ij}. For, substituting for $p_0(s)$ its expression as given by Wishart [8] we shall have

$$K \int_{W_1} g(s) p_0(s) \, ds \equiv \int_{W_1} g(s) \mid S \mid^{\frac{1}{2}(f-k)} \exp\left[-\frac{1}{2} \sum_{i,j=1}^{k} c_{ij} s_{ij} \right] ds \equiv 0 \tag{18}$$

where $| S | = | s_{ij} |$ and K is some constant. Differentiating (18) successively with respect to the c_{ij} and dividing the results by K, we shall regain the equations (16). Hence it follows that $g(s) \equiv 0$.

This being established, let w be any region belonging to D and rewrite the equation (14), so that

$$(\sqrt{2\pi})^{-k(f+1)} C^{\frac{1}{2}(f+1)} \int_w \exp\left[-\frac{1}{2} \sum_{i,j=1}^{k} c_{ij} \{ (z_i - \zeta_i)(z_j - \zeta_j) + v'_{ij} \} \right] d(z, z') = \epsilon + \frac{\alpha}{2} \sum_{i,j=1}^{k} c_{ij} \zeta_i \zeta_j + R \tag{19}$$

Setting all the ζ_i to zero in both sides of (19), we have

$$\int_w p_0(z, z') \, d(z, z') \equiv \epsilon \tag{20}$$

identically in the c_{ij}. Differentiating (19) once with respect to ζ_i and afterwards setting all the ζ_i to zero, we easily get

$$\int_w z_i p_0(z, z') \, d(z, z') \equiv 0 \qquad (i = 1, 2, \cdots k) \tag{21}$$

for all possible values of the c_{ij}.

Finally, differentiating (19) with respect to ζ_i and then to ζ_j and putting all $\zeta_i = 0$ in the result we obtain

$$\int_w \left\{ \left(\sum_{h=1}^{k} c_{ih} z_h \right) \left(\sum_{h=1}^{k} c_{jh} z_h \right) - c_{ij} \right\} p_0(z, z') \, d(z, z') \equiv \alpha c_{ij} \qquad (i, j = 1, 2, \cdots k)$$

whence, renumbering (20)

$$\sum_{h,l=1}^{k} c_{ih} c_{jl} q_{hl} \equiv \beta c_{ij} \qquad (i, j = 1, 2, \cdots, k) \tag{22}$$

in which we denote by $\beta = \alpha + \epsilon$ and

$$q_{hl} = \int_w z_h z_l p_0(z, z')\, d(z, z') \qquad (h, l = 1, 2, \cdots, k)$$

If we denote by Q the matrix of order k formed of the elements q_{hl}, we see that (22) may be written as

$$CQC \equiv \beta C,$$

whence, since C has its inverse matrix, C^{-1},

$$Q \equiv \beta C^{-1}$$

i.e.,

$$q_{ij} \equiv \beta c_{ij}^{(-1)} \qquad (i, j = 1, 2, \cdots, k) \tag{23}$$

where $c_{ij}^{(-1)}$ denotes the element in the matrix C^{-1} which corresponds to the element c_{ij} in the matrix C.

Conditions (20), (21) and (23) are necessary for the region w to belong to the class D. They are evidently also sufficient.

Let us evaluate the integrals in (20), (21) and the q_{ij} by first evaluating the surface integrals on any surface, say $G(s)$, on which all the s_{ij} have constant values, and then integrating the results with respect to the s_{ij} over a region, say w_1, of the s_{ij} contained in W_1. Thus we may write (20), (21) and (23) in the form

$$\int_{w_1} f(s)p_0(s)\, ds \equiv \epsilon, \qquad \int_{w_1} g_i(s)p_0(s)\, ds \equiv 0, \qquad \int_{w_1} \varphi_{ij}(s)p_0(s) \equiv \beta c_{ij}^{(-1)}, \qquad (i, j = 1, 2, \cdots, k), \tag{24}$$

where

$$f(s) = \frac{1}{p_0(s)} \int_{G(s)} p_0(z, z')\, dG(s)$$

$$g_i(s) = \frac{1}{p_0(s)} \int_{G(s)} z_i p_0(z, z')\, dG(s)$$

$$\varphi_{ij}(s) = \frac{1}{p_0(s)} \int_{G(s)} z_i z_j p_0(z, z')\, dG(s)$$

It is readily verified that the function $p_0(z, z')/p_0(s)$ is free from the parameters c_{ij}, and consequently so are the functions $f(s)$, $g_i(s)$, $\varphi_{ij}(s)$. Besides, we can extend the definition of these functions in the whole domain W_1 by assigning them the value zero outside of the region w_1. Doing this we can now write the equations (24) as

$$\int_{W_1} (f(s) - \epsilon)p_0(s)\, ds \equiv 0, \qquad \int_{W_1} g_i(s)p_0(s)\, ds \equiv 0,$$

$$\int_{W_1} [\varphi_{ij}(s) - \gamma s_{ij}]\, p_0(s)\, ds \equiv 0 \qquad (i, j = 1, 2, \cdots, k) \tag{25}$$

where $\gamma = \dfrac{1}{f + 1}\, \beta$.

Now all the equations (25) are of the form (17); consequently, according to the already established result and remembering the definitions of the functions $f(s)$, $g_i(s)$ and $\varphi_{ij}(s)$, we must have

$$\int_{G(s)} p_0(z, z')\, dG(s) = \epsilon p_0(s) \tag{26}$$

$$\int_{G(s)} z_i p_0(z, z')\, dG(s) = 0 \tag{27}$$

$$\int_{G(s)} z_i z_j p_0(z, z')\, dG(s) = \gamma s_{ij} p_0(s) \tag{28}$$

in the whole domain W_1.

Hence the most general region belonging to the class D is constructed as follows. On any surface $s_{ij} = \text{const.}$ $(i, j = 1, 2, \cdots k)$ we take an areal region such that it satisfies the equations (26)–(28); we then allow the s_{ij} to vary in the whole domain W_1. Equations (28) may now be replaced by

$$\int_{G(s)} \left(\frac{z_1^2}{s_{11}} - \frac{z_i z_j}{s_{ij}}\right) p_0(z, z') = 0, \qquad (i, j = 1, 2, \cdots, k) \tag{28'}$$

Let us call w_0 the region defined by $T \geq T_\epsilon$. Since w_0 belongs to the class D (cf. (12)), its cross section, say $G_0(s)$, by any surface $s_{ij} = \text{const.}$ $(i, j = 1, 2, \cdots, k)$ must satisfy the equations (26), (27) and (28′). Since $\gamma = \dfrac{1}{f+1}(\alpha + \epsilon)$, all we have to prove now is that among all the areal regions $G(s)$ satisfying the equations (26), (27) and (28′) it is the region $G_0(s)$ that gives the largest possible value to $\gamma p_0(s)$. Now

$$\gamma p_0(s) = \int_{G(s)} \frac{z_1^2}{s_{11}} p_0(z, z')\, d(z, z') \tag{29}$$

and, according to a Lemma of Neyman and Pearson, [3, p. 10] the right-hand side of (29) will attain its maximum value if $G(s)$ is defined by an inequality of the form

$$\frac{z_1^2}{s_{11}} \geq \sum_{i,j=1}^{k} a_{ij}\left(\frac{z_1^2}{s_{11}} - \frac{z_i z_j}{s_{ij}}\right) + \sum_{i=1}^{k} b_i z_i + c \tag{30}$$

where the a_{ij}, b_i and c are constants so determined as to enable the region $G(s)$ to satisfy the equations (26)–(28). We shall show presently that the region $G_0(s)$ is defined by such an inequality.

The inequality $T \geq T_\epsilon$ may be written as

$$\frac{|v'_{ij}|}{|v'_{ij} + z_i z_j|} \leq \frac{1}{1 + T_\epsilon}$$

f.e.

$$\frac{|s_{ij} - z_i z_j|}{|s_{ij}|} \leq \frac{1}{1 + T_\epsilon},$$

or

$$\sum_{i,j=1}^{k} s_{ij}^{(-1)} z_i z_j \geq \frac{T_\epsilon}{1+T_\epsilon} \tag{31}$$

where $s_{ij}^{(-1)}$ denotes the (i, j)th element in the inverse matrix of $\| s_{ij} \|$. The region $G_0(s)$ is therefore defined by the same inequality (31) in which we regard the s_{ij} as constants.

If we put

$$a_{ij} = \frac{1}{k} s_{ij} s_{ij}^{(-1)}, \qquad b_i = 0, \qquad c = \frac{1}{k} \frac{1}{1+T_\epsilon} \qquad (i, j = 1, 2, \cdots, k)$$

in (30) we can easily reduce the inequality (30) into (31).

The proof is now complete.

3. **Note on Applications of T.** It is already known that the T-test may be used for the following purposes (a) and (b):

(a) Given a k-variate normal surface

$$p(x) = (\sqrt{2\pi})^{-k} C^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}(x_i - \xi_i)(x_j - \xi_j)\right]$$

with the unknown ξ_i and c_{ij}. n observations

$$(x_{1l}, x_{2l}, \cdots, x_{kl}), \qquad (l = 1, 2, \cdots, n)$$

having been made, it is required to test the hypothesis that the ξ_i have the particular values ξ_i^0 for $i = 1, 2, \cdots, k$.

Here we use the T-test with

$$\left.\begin{aligned} z_i = \sqrt{n}(\bar{x}_i - \xi_i^0), \qquad v'_{ij} = \sum_{l=1}^{m} (x_l - \bar{x}_i)(\bar{x}_{jl} - \bar{x}_j) \\ \zeta_i = \sqrt{n}\,(\xi_i - \xi_i^0), \qquad f = n - 1 \\ \text{where} \qquad \bar{x}_i = \frac{1}{n}\sum_{l=1}^{n} x_{il} \end{aligned}\right\} (i, j = 1, 2, \cdots, k)$$

(b) Given two k-variate normal surfaces

$$p_1(x) = (\sqrt{2\pi})^{-k} C^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}(x_i - \xi_i)(x_j - \xi_j)\right)$$

$$p_2(x) = (\sqrt{2\pi})^{-k} C^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}(y_i - \eta_i)(y_j - \eta_j)\right)$$

where the c_{ij} are common to the two surfaces while all the ξ_i, ξ_j, c_{ij} are unknown. Samples of n_1 and n_2 having been drawn respectively from the two populations, to test the hypothesis that $\xi_i = \eta_i$ for all i.

Let the samples be

$$(x_{1l}, x_{2l}, \cdots, x_{kl}), \qquad (l = 1, 2, \cdots, n_1)$$

and

$$(y_{1h}, y_{2h}, \cdots, y_{kh}), \qquad (h = 1, 2, \cdots, n_2)$$

Let

$$\bar{x}_i = \frac{1}{n_1}\sum_{i=1}^{n_1} x_{il}, \qquad \bar{y}_i = \frac{1}{n_2}\sum_{h=1}^{n_2} y_{ih} \qquad (i = 1, 2, \cdots, k)$$

We use the T-test with

$$z_i = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(\bar{x}_i - \bar{y}_i), \qquad v'_{ij} = \sum_{l=1}^{n_1}(x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j) + \sum_{h=1}^{n_2}(y_{ih} - \bar{y}_i)(y_{ih} - \bar{y}_j)$$

$$\zeta_i = \sqrt{\frac{n_1 n_2}{n_1 + n_2}}\,(\xi_i - \eta_i), \qquad f = n_1 + n_2 - 2$$

$$(i, j = 1, 2, \cdots, k)$$

A third application of T, which appears to be novel, is the following:

(c) Given a $(k + 1)$-variate normal surface

$$p(x) = (\sqrt{2\pi})^{-(k+1)} D^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k+1} d_{ij}(x_i - \xi_i)(x_j - \xi_j)\right], \qquad D = |\,d_{ij}\,|,$$

where the ξ_i and d_{ij} are all unknown. n observations

$$(x_{1l}, x_{2l}, \cdots, x_{k+1,l}) \qquad (l = 1, 2, \cdots, n)$$

having been made, to test the hypothesis that all the ξ_i are equal.

If we put

$$y_i = x_i - x_{k+1} \qquad (i = 1, 2, \cdots, k),$$

then we have a k-variate normal surface for the variables y_i.

$$p(y) = (\sqrt{2\pi})^{-k} C^{\frac{1}{2}} \exp\left[-\frac{1}{2}\sum_{i,j=1}^{k} c_{ij}(y_i - \eta_i)\right]$$

where $\eta_i = \xi_i - \xi_{k+1}$ $(i = 1, 2, \cdots, k)$. Thus the problem is reduced to testing the hypothesis that $\eta_i = 0$ for $i = 1, 2, \cdots, k$ and therefore belongs to the type (a). Write

$$y_{il} = x_{il} - x_{k+1,l} \qquad (i = 1, 2, \cdots, k; l = 1, 2, \cdots, n)$$

and

$$\bar{y}_i = \frac{1}{2}\sum_{l=1}^{n} y_{il}, \qquad (i = 1, 2, \cdots, k).$$

We use the T-test with

$$\left.\begin{aligned} z_i = \sqrt{n}\,\bar{y}_i, \qquad v'_{ij} = \sum_{l=1}^{n} (y_{il} - \bar{y})(y_{jl} - \bar{y}_j) \\ \zeta_i = \sqrt{n}\,\eta_i, \qquad f = n + 1 \end{aligned}\right\} (i, j = 1, 2, \cdots, k)$$

Although there are no simple expressions for the c_{ij}, there is one for the parameter $\Sigma\, c_{ij}\eta_i\eta_j$, on which alone the distribution of T depends. We have indeed

$$\sum_{i,j=1}^{k} c_{ij}\eta_i\eta_j = \frac{1}{D}\begin{vmatrix} \sigma_{11} & \cdots & \sigma_{1,k+1} & \xi_1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{k+1,1} & \cdots & \sigma_{k+1,k+1} & \xi_{k+1} & 1 \\ \xi_1 & \cdots & \xi_{k+1} & 0 & 0 \\ 1 & \cdots & 1 & 0 & 0 \end{vmatrix}$$

where

$$D = \begin{vmatrix} \sigma_{11} & \cdots & \sigma_{1,k+1} & 1 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{k+1,1} & \cdots & \sigma_{k+1,k+1} & 1 \\ 1 & \cdots & 1 & 0 \end{vmatrix}$$

where σ_{ij} is the covariance between x_i and x_j.

Expressing T in terms of the original variables x, we have

$$T = -\frac{1}{D'}\begin{vmatrix} s_{11}\ s_{12} & \cdots & s_{1,k+1} & \bar{x}_1 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{k+1,1} & \cdots & s_{k+1,k+1} & \bar{x}_{k+1} & 1 \\ \bar{x}_1 & \cdots & \bar{x}_{k+1} & 0 & 0 \\ 1 & \cdots & 1 & 0 & 0 \end{vmatrix}$$

where

$$D' = \begin{vmatrix} s_{11} & \cdots & s_{1,k+1} & 1 \\ \cdots & \cdots & \cdots & \cdots \\ s_{k+1,1} & \cdots & s_{k+1,k+1} & 1 \\ 1 & \cdots & 1 & 0 \end{vmatrix}$$

and where

$$\bar{x}_i = \frac{1}{n}\sum_{l=1}^{n} x_{il}, \qquad s_{ij} = \frac{1}{n}\sum_{l=1}^{n} (x_{il} - \bar{x}_i)(x_{jl} - \bar{x}_j), \qquad (i, j = 1, 2, \cdots, k + 1)$$

Therefore T is independent of which variable has been taken as the $(k + 1)$st.

UNIVERSITY COLLEGE, LONDON.

REFERENCES

[1] H. HOTELLING, *Ann. Math. Statist.*, Vol. 2 (1931) pp. 359-378.

[2] J. NEYMAN, *Bull. Soc. Math. France*, Vol. 63 (1935) pp. 246–266.
[3] J. NEYMAN AND E. S. PEARSON, *Statist. Res. Mem.*, Vol. 1 (1936) pp. 1–37.
[4] J. NEYMAN AND E. S. PEARSON, *Statist. Res. Mem.*, Vol. 2 (1938).
[5] P. C. TANG, *Statist. Res. Mem.*, Vol. 2 (1938).
[6] E. C. TITCHMARSH, *Introduction to the Theory of Fourier Integrals*. Oxford Univ. Press (1937).
[7] S. S. WILKS, *Biometrika*, Vol. 24 (1932) pp. 471–494.
[8] J. WISHART, *Biometrika*, Vol. 20A (1928) pp. 32–52.

GENERALIZATION OF THE INEQUALITY OF MARKOFF

BY A. WALD

1. **Introduction.** Denote by X a random variable and by M_r the expected value $E \mid X - x_0 \mid^r$ of $\mid X - x_0 \mid^r$ for any integer r where x_0 denotes a given real value. M_r is also called the absolute moment of order r about the point x_0. For any positive number d, denote by $P(-d < X - x_0 < d)$ the probability that $\mid X - x_0 \mid < d$. The inequality of Markoff can be written as follows

$$P(-d < X - x_0 < d) \geq 1 - \frac{M_r}{d^r} \tag{1}$$

The inequality (1) is also called, for $r = 2$, the inequality of Tchebyscheff. The inequality (1) can be written in the following way:

$$P(-\xi\sqrt[r]{M_r} < X - x_0 < \xi\sqrt[r]{M_r}) \geq 1 - \frac{1}{\xi^r}.$$

Substituting in the above inequality s for r and $\bar{\xi}\,\dfrac{\sqrt[r]{M_r}}{\sqrt[s]{M_s}}$ for ξ we get

$$P(-\bar{\xi}\sqrt[r]{M_r} < X - x_0 < \bar{\xi}\sqrt[r]{M_r}) \geq 1 - \frac{1}{\bar{\xi}^s}\left(\frac{\sqrt[s]{M_s}}{\sqrt[r]{M_r}}\right)^s, \tag{2}$$

where r and s denote any integers and $\bar{\xi}$ denotes an arbitrary positive value.[1] Substituting in (2) $2k$ for s and 2 for r, we get the inequality of K. Pearson.[2] By other substitutions we get the formulae of Lurquin, Cantelli, etc.[3]

As is well known, the inequality (1) cannot be improved[4] for $d \geq \sqrt[r]{M_r}$. That is to say, to every $\epsilon > 0$ a random variable Y can be given such that

$$E \mid Y - x_0 \mid^r = E \mid X - x_0 \mid^r \quad \text{and} \quad P(-d < Y - x_0 < d) < 1 - \frac{M_r}{d^r} + \epsilon.$$

If the absolute moments $M_{i_1} = E(\mid X - x_0 \mid^{i_1}), \cdots, M_{i_j} = E \mid X - x_0 \mid^{i_j}$ of a random variable X are given (and no further data about X are known), then we shall say that a_d is the "*sharp*" *lower limit* of $P(-d < X - x_0 < d)$ if the following two conditions are fulfilled:

(1) For each random variable Y, for which $E \mid Y - x_0 \mid^{i_1} = E \mid X - x_0 \mid^{i_1}, \cdots, E \mid Y - x_0 \mid^{i_j} = E \mid X - x_0 \mid^{i_j}$, the inequality $P(-d < Y - x_0 < d) \geq a_d$ holds.

[1] The formula (2) has been given by A. Guldberg, *Comptes Rendus*, Paris, Vol. 175, p. 679.

[2] *Biometrika*, Vol. XII (1918–1919) pp. 284–296.

[3] E. Lurquin, *Comptes Rendus*, Paris, Vol. 175, p. 681. Also Cantelli, *Rendinconti delle Reale Academia dei Lincei*, 1916.

[4] See for instance, R. v. Mises, *Wahrscheinlichkeitsrechnung*, Leipzig, Vienna, Deuticke, 1931, p. 36.

(2) To each $\epsilon > 0$, a random variable Y can be given such that $E \mid Y - x_0 \mid^{i_\nu} = E \mid X - x_0 \mid^{i_\nu}$ $(\nu = 1, \cdots, j)$ and $P(-d < Y - x_0 < d) < a_d + \epsilon$.

In other words, a_d is the *limes inferior*[5] of the probabilities $P(-d < Y - x_0 < d)$ formed for all random variables Y for which the i_ν-th absolute moment about the point x_0 is equal to the i_ν-th moment of X about the point x_0 $(\nu = 1, \cdots, j)$.

PROBLEM: *The absolute moments $M_{i_1}, M_{i_2}, \cdots, M_{i_j}$ of a random variable X are given about the point x_0, where $i_1, i_2, \cdots, i_j$ denote any integers and M_{i_ν} denotes the moment of order i_ν $(\nu = 1 \cdots k)$. It is required to calculate the "sharp" lower limit of the probability $P(-d < X - x_0 < d)$ for any positive value d.*

If only a single moment M_r is given, our problem is already solved, because the inequality (1) gives us the "sharp" lower limit for $d \geq \sqrt[r]{M_r}$ and for $d < \sqrt[r]{M_r}$ the "sharp" limit is obviously equal to zero. But the case in which even two moments M_r and M_s are given has not yet been solved. The formula (2) gives us a limit for $P(-d < X - x_0 < d)$, but this limit is not "sharp," as can easily be shown.

We shall give here some results concerning the general case, and the complete solution if only two moments M_r and M_s are given. We shall see that the "sharp" limit is considerably greater than the limit given by (2).

2. Some Propositions Concerning the General Case. We shall call a random variable X non-negative if $P(X < 0) = 0$. Since the absolute moments of the non-negative random variable $Y = \mid X - x_0 \mid$ about the origin are equal to the absolute moments of X about the point x_0 and since $P(Y < d) = P(-d < X - x_0 < d)$, the following proposition holds true:

(I) *Denote by $M_{i_1}, \cdots, M_{i_j}$ the absolute moments of order $i_1, \cdots, i_j$ of a certain random variable X about the point x_0. The limes inferior of the probabilities $P(-d < Y - x_0 < d)$ is equal to the limes inferior of the probabilities $P(Z < d)$, where $P(-d < Y - x_0 < d)$ is formed for all random variables Y for which the i_ν-th absolute moment about x_0 is equal to M_{i_ν} $(\nu = 1, \cdots, j)$, and $P(Z < d)$ is formed for all non-negative random variables Z for which the i_ν-th moment about the origin is equal to M_{i_ν} $(\nu = 1, \cdots, j)$.*

On account of the proposition (I) we can restrict ourselves to the consideration of non-negative random variables and of the moments about the origin.

A random variable X for which k different values $x_1, \cdots, x_k$ exist such that the probability $p(x_i)$ of x_i $(i = 1, \cdots, k)$ is positive and $\sum_{i=1}^{k} p(x_i) = 1$, is called an *arithmetic* random variable of degree k. A random variable X will be called t-limited, if $P(-t \leq X \leq t) = 1$ We shall prove the following proposition.

(II). *Let us denote by $M_{i_1}, M_{i_2}, \cdots, M_{i_j}$ the absolute moments of order $i_1, \cdots, i_j$ of a certain non-negative random variable X, about the origin. Denote by $\Omega(k, t)$ the set of all non-negative t-limited arithmetic random variables of*

[5] The *limes inferior* of a set N of numbers is the greatest value y for which the inequality $y \leq x$ for each element x of N holds true. This is also called greatest lower bound.

degree $\leq k$, *for which the* i_ν*th moment about the origin is equal to* M_{i_ν} $(\nu = 1, \cdots, j)$. $\Omega(k, t)$ *is supposed to be not empty.*[6] *Denote further by* $a(d, k, t)$ *the limes inferior of the probabilities* $P(Y < d)$ *formed for all random variables* Y *of the set* $\Omega(k, t)$. *Then we can say: There exists in* $\Omega(k, t)$ *a random variable* Z *for which* $P(Z < d) = a(d, k, t)$. *If* $0 < a(d, k, t) < 1$ *and* Z *is a random variable in* $\Omega(k, t)$ *for which* $P(Z < d) = a(d, k, t)$, *then there exist at most* $j - 1$ *different positive values* $x_1, \cdots, x_{j-1}$ *such that* $x_i \neq d$, $x_i \neq t$ *and the probability* $p(x_i)$ *of* x_i *is positive* $(i = 1, 2, \cdots, j - 1)$.

At first we shall prove that there exists a random variable Z in $\Omega(k, t)$ such that $P(Z < d) = a(d, k, t)$. Since $a(d, k, t)$ is the *limes inferior* of $P(Y < d)$ formed for all random variables Y in $\Omega(k, t)$, there exists in $\Omega(k, t)$ a sequence $\{Z_i\}$ $(i = 1, 2, \cdots$, ad inf.) of random variables, such that $\lim_{i=\infty} P(Z_i < d) = a(d, k, t)$. Arranged in ascending order of magnitude, the values of Z_i which have a positive probability are denoted by $x_{i,1}, x_{i,2}, \cdots, x_{i,k_i}$. Since Z_i is a t-limited non-negative arithmetic random variable of degree $\leq k$, we have $k_i \leq k$ and $0 \leq x_{i,r} \leq t$ $(r = 1, \cdots, k_i)$. It follows easily from this fact that there exists a subsequence $\{Z_{i_\nu}\}$ $(\nu = 1, 2, \cdots$, ad inf.) of $\{Z_i\}$ with two properties: First, that the variables $Z_{i_1}, Z_{i_2}, \cdots$ are of the same degree (say s), that is to say $k_{i_\nu} = s$ $(\nu = 1, 2, \cdots$, ad inf.); and second, that the sequence $\{x_{i_\nu,r}\}$ $(\nu = 1, 2, \cdots$, ad inf.) converges for each integer $r \leq s$. Let us denote $\lim x_{i_\nu,r}$ by x_r $(r = 1, 2, \cdots, s)$, and the probability that Z_{i_ν} takes the value $x_{i_\nu,r}$ by $p_{i_\nu,r}$. It is obvious that there exists a subsequence $\{Z_{n_\nu}\}$ $(\nu = 1, 2, \cdots$ ad inf.) of $\{Z_{i_\nu}\}$ such that the sequence $\{p_{n_\nu,r}\}$ converges with increasing ν. Let us denote $\lim_{\nu=\infty} p_{n_\nu,r}$ by p_r, $(r = 1, 2, \cdots s)$. Since $p_{i_\nu,1} + \cdots + p_{i_\nu,r} = 1$, $\sum_{r=1}^{s} p_r = 1$ must hold true. We consider the random variable Z for which the probability that $Z = x_r$ is equal to p_r $(r = 1, 2, \cdots s)$ and for which no values except $x_1, \cdots, x_r$ are possible. The random variable Z is obviously an element of $\Omega(k, t)$ and $P(Z < d) = a(d, k, t)$.

Let us consider the case in which $0 < a(d, k, t) < 1$ and denote by Z a random variable of $\Omega(k, t)$ for which $P(Z < d) = a(d, k, t)$. We shall prove that there exist at most $j - 1$ different positive values $x_1, \cdots, x_{j-1}$ such that $x_i \neq d$, $x_i \neq t$ and the probability $p(x_i)$ of x_i is positive $(i = 1, 2, \cdots, j - 1)$. In order to prove this statement we shall suppose that there exist j different positive points $x_1, \cdots, x_j$ such that $x_i \neq d$, $x_i \neq t$ and $p(x_i) > 0$ $(i = 1, 2, \cdots, j)$. Then we can write

$$\sum_{\nu=1}^{j} x_\nu^{i_1} p(x_\nu) = M_{i_1} - \sum x^{i_1} p(x)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_{\nu=1}^{j} x_\nu^{i_j} p(x_\nu) = M_{i_j} - \sum x^{i_j} p(x),$$

[6] This is certainly the case, if we choose k and t great enough.

where the summations on the right hand sides are to be taken over all values of x which are different from $x_1, \cdots, x_j$ and for which $p(x) > 0$.

Since $P(Z < d) = a(d, k, t)$ and $0 < a(d, k, t) < 1$ by hypothesis, there exist two non-negative values b and c, such that $b < d$, $c \geq d$, $p(b) > 0$, and $p(c) > 0$.

We define a new arithmetic random variable Z' as follows: $p'(b) = p(b) - \epsilon$, $p'(c) = p(c) + \epsilon$, and for all other values $p'(x) = p(x)$, where $p'(x)$ denotes the probability that $Z' = x$, and ϵ a positive number less than $p(b)$. Z' is obviously a non-negative arithmetic variable of the same degree as Z. The moments about the origin of the order $i_1, i_2, \cdots, i_j$ of Z' will in general not be equal to the corresponding moments of Z. However this can be obtained by a small displacement of the points $x_1, \cdots, x_j$ into a system of neighboring points $\bar{x}, \cdots, \bar{x}_j$, provided that ϵ is small enough. In order to show this, we have only to prove that the functional determinant

$$\Delta = \begin{vmatrix} i_1\bar{x}_1^{i_1-1}, & \cdots, & i_1\bar{x}_j^{i_1-1} \\ i_2\bar{x}_1^{i_2-1}, & \cdots, & i_2\bar{x}_j^{i_2-1} \\ \cdots & \cdots & \cdots \\ i_j\bar{x}_1^{i_j-1}, & \cdots, & i_j\bar{x}_j^{i_j-1} \end{vmatrix} p'(x_1) \cdots p'(x_j)$$

of the functions $f_1(\bar{x}_1, \cdots, \bar{x}_j) = \sum_{\nu=1}^{j} \bar{x}_\nu^{i_1} p'(x_\nu), \cdots, f_j(\bar{x}_1, \cdots \bar{x}_j) = \sum_{\nu=1}^{j} \bar{x}_\nu^{i_1} p'(x_\nu)$ does not vanish at the point $\bar{x}_1 = x_1, \cdots, \bar{x}_j = x_j$. Since $p'(x_1), p'(x_2), \cdots p'(x_j)$ are not equal to zero, we have only to show that

$$\Delta^* = \begin{vmatrix} x_1^{i_1-1}, & \cdots, & x_j^{i_1-1} \\ \cdots & \cdots & \cdots \\ x_1^{i_j-1}, & \cdots, & x_j^{i_j-1} \end{vmatrix} = \begin{vmatrix} 1, & \cdots, & 1 \\ x_1^{i_2-i_1}, & \cdots, & x_j^{i_2-i_1}, \\ \cdots & \cdots & \cdots \\ x_1^{i_j-i_1}, & \cdots & x_j^{i_j-i_1} \end{vmatrix} x_1^{i_1-1} \cdots x_j^{i_1-1} \neq 0$$

where $i_2 - i_1, \cdots, i_j - i_1$ can be assumed positive by denoting by i_1 the smallest of the integers $i_1, i_2, \cdots, i_j$.

Let us consider the polynomial in x given by

$$R(x) = \begin{vmatrix} 1, & \cdots, & 1, & 1 \\ x_1^{i_2-i_1}, & \cdots & x_{j-1}^{i_2-i_1}, & x^{i_2-i_1} \\ \cdots & \cdots & \cdots & \cdots \\ x_1^{i_j-i_1}, & \cdots, & x_{j-1}^{i_j-i_1}, & x^{i_j-i_1} \end{vmatrix}$$

According to a well-known algebraic proposition, the number of positive roots of $R(x)$ is less than or equal to the number of changes of sign in the sequence of coefficients of $R(x)$. Since the number of changes of sign in $R(x)$ is obviously less than or equal to $j - 1$, the number of positive roots of $R(x)$ is also less than or equal to $j - 1$. On the other hand $x = x_1, \cdots, x = x_{j-1}$ are $j - 1$ positive roots of $R(x)$. Hence for any positive value $x \neq x_1, \neq x_2, \cdots,$

$\neq x_{j-1}$ the polynomial $R(x)$ does not vanish. Thus $R(x_j)$ and therefore also Δ^* and Δ are not equal to zero.

Let us denote by Z^* the random variable which we get from Z' by a small displacement of the points $x_1, \cdots, x_j$ into a system of neighboring points $\bar{x}_1, \cdots, \bar{x}_j$, such that the moment of order i_ν of Z^* about the origin becomes equal to M_{i_ν} $(\nu = 1, 2, \cdots, j)$. By choosing ϵ small enough we can obtain the values $\bar{x}_1, \cdots, \bar{x}_j$ as near to $x_1, \cdots, x_j$ as we like. In particular, ϵ can be chosen so small that $\bar{x}_1, \cdots, \bar{x}_j$ are positive numbers less than t, and $\bar{x}_i > d$ or $< d$ accordingly as $x_i >$ or $< d$. Then Z^* is obviously an element of $\Omega(k, t)$. But for Z^*

$$P(Z^* < d) = P(Z' < d) = P(Z < d) - \epsilon = a(d, k, t) - \epsilon$$

holds true, which is a contradiction because $a(d, k, t)$ is the *limes inferior* of $P(Y < d)$ formed for all random variables Y contained in $\Omega(k, t)$. Hence our assumption that there exist j different positive numbers $x_1, \cdots, x_j$, for which $x_i \neq d$, $x_i \neq t$ and $p(x_i) > 0$ $(i = 1, 2, \cdots, j)$, cannot be true, and the proposition II is proved in all its parts.

It follows from the proposition II that $a(d, k, t)$ is independent of k. On account of this fact and of the fact that any random variable X can be arbitrarily well approximated by arithmetic random variables, we get the proposition:

III. *Let us denote by $M_{i_1}, \cdots, M_{i_j}$ the moments about the origin of order $i_1, \cdots, i_j$ of a certain non-negative random variable. Denote by $\Omega(t)$ the set of all non-negative t-limited random variables, for which the i_ν-th moment about the origin is equal to M_{i_ν} $(\nu = 1, \cdots, j)$. Denote further by $a(d, t)$ the limes inferior of the probabilities $P(Y < d)$ formed for all random variables Y contained in $\Omega(t)$. Then we can say: There exists in $\Omega(t)$ a random variable Z for which $P(Z < d) = a(d, t)$. If $0 < a(d, t) < 1$ and Z is a random variable for which $P(Z < d) = a(d, t)$, then there exist at most $j - 1$ different positive numbers $x_1, \cdots, x_{j-1}$, such that $x_i \neq d$, $x_i \neq t$, and the probability that $Z = x_i$, is positive* $(i = 1, 2, \cdots, j - 1)$

It is obvious that $a(d, t)$ decreases monotonically with increasing t. Hence $\lim_{t=\infty} a(d, t)$ exists and it can be easily shown that:

$a(d, t)$ converges towards a_d if $t \to \infty$.

3. **Solution of the Problem if Only Two Moments are Given.** Let us denote by M_r and M_s the absolute moments respectively of order r and s about the point x_0 of a certain random variable X, where r and s $(r < s)$ denote any integers.

Let us first consider the case

$$(\alpha) \qquad \frac{M_r}{d^r} \leq \frac{M_s}{d^s}$$

It follows from (1) that

$$a_d \geq 1 - \frac{M_r}{d^r}$$

We shall show that $a_d = 1 - \frac{M_r}{d^r}$ if $\frac{M_r}{d^r} \leq 1$. For this purpose let us consider the arithmetic random variable Y_b of degree 3 defined as follows:

$$p(x_0 + d) = \frac{M_r}{d^r} - \frac{\epsilon}{2}, \qquad p(x_0 + d + b) = \frac{\epsilon}{2}\left(\frac{d}{d+b}\right)^r$$

$$p(x_0) = 1 - p(x_0 + d) - p(x_0 + d + b)$$

where ϵ is a positive number and $p(u)$ denotes the probability for $Y_b = u$. The r-th moment about x_0 of Y_b is obviously equal to M_r. On account of (α) the s-th moment of Y_b about x_0 is less than or equal to M_s for $b = 0$. On the other hand the s-th moment of Y_b about x_0 will be greater than M_s if b is sufficiently large. Hence there exists a non-negative value b_0 such that the s-th moment of Y_{b_0} is equal to M_s.

Since $P(-d < Y_{b_0} - x_0 < d) = 1 - \frac{M_r}{d^r} + \frac{\epsilon}{2} - \frac{\epsilon}{2}\left(\frac{d}{d+b_0}\right)^r < 1 - \frac{M_r}{d^r} + \frac{\epsilon}{2}$ and since ϵ can be chosen arbitrarily small, we have

$$a_d = 1 - \frac{M_r}{d^r}.$$

If $\frac{M_r}{d^r} \geq 1$, then a_d is equal to zero, because a_d decreases monotonically with decreasing d and $a_d = 0$ for $d = \sqrt[r]{M_r}$.

We have now to consider the case

$$(\beta) \qquad \frac{M_r}{d^r} > \frac{M_s}{d^s}$$

First we shall show that

$$(3) \qquad \frac{M_r}{d^r} < 1.$$

In fact, if $\frac{M_r}{d^r}$ were ≥ 1, then making use of (β) we have $\left(\frac{M_r}{d^r}\right)^{\frac{s}{r}} \geq \frac{M_r}{d^r} > \frac{M_s}{d^s}$, and hence $(M_r)^{\frac{s}{r}} > M_s$. But this is not possible, because according to the well-known inequalities between moments, $(M_r)^{\frac{s}{r}}$ is less than or equal to M_s. It follows from (3) and (β) that

$$(4) \qquad \frac{M_s}{d^s} < 1.$$

In order to calculate a_d, we shall apply the propositions found in section 2. On account of the proposition I, a_d is equal to the *limes inferior* of $P(Y < d)$

where $P(Y < d)$ is formed for all non-negative random variables Y for which the r-th moment about the origin is equal to M_r and the s-th moment about the origin is equal to M_s. Hence we can restrict ourselves to the consideration of non-negative random variables and of the moments about the origin.

We shall show that $0 < a(d, t)$ holds for any positive value t. In order to prove this, it is sufficient to show that $a_d > 0$ since $a(d, t) \geq a_d$. It follows from the inequality (1) that $a_d \geq 1 - \frac{M_r}{d^r}$. Since, according to (3), $\frac{M_r}{t^r} < 1$, we have $a_d > 0$, and therefore also

$$a(d, t) > 0 \tag{5}$$

Let us see whether $a(d, t) < 1$. If $M_s = (M_r)^{\frac{s}{r}}$, then, as is well-known, only a single non-negative random variable X exists for which the r-th moment about the origin is equal to M_r and the s-th moment is equal to $(M_r)^{\frac{s}{r}}$, namely the arithmetic random variable X of degree 1 for which the probability that $X = \sqrt[r]{M_r}$ is equal to 1. Since $\sqrt[r]{M_r} < d$, as can be seen from (3), we have $P(X < d) = 1$, and therefore $a_d = 1$. Hence in this case our problem is already solved and we have to consider only the alternative:

$$M_s = M_r^{\frac{s}{r}} + \sigma^2 \; (\sigma^2 > 0) \tag{6}$$

We shall show that $a(d, t) < 1$ for $t > \sqrt[r]{M_r} + d_r$. For this purpose let us consider the non-negative arithmetic random variable Y_ϵ of the degree 3 defined as follows:

$$p(\sqrt[r]{M_r}) = 1 - \epsilon, \; p(t) = \epsilon \frac{M_r}{t^r} < \epsilon \frac{M_r}{t^r} < \epsilon$$

$$p(0) = 1 - p(\sqrt[r]{M_r}) - p(t) = \epsilon - \epsilon \frac{M_r}{t^r},$$

where $p(u)$ denotes the probability for $Y_\epsilon = u$, and ϵ is a positive number < 1.

The r-th moment of Y_ϵ is equal to

$$M_r p(\sqrt[r]{M_r}) + t^r p(t) = M_r.$$

The s-th moment of Y_ϵ is given by the expression

$$A = M_r^{\frac{s}{r}} p(\sqrt[r]{M_r}) + t^s p(t) = (1 - \epsilon) M_r^{\frac{s}{r}} + \epsilon t^s \frac{M_r}{t^r}.$$

On account of (6), A is less than M_s for $\epsilon = 0$. For $\epsilon = 1$ we have

$$A = t^{s-r} M_r > d^{s-r} M_r.$$

Since from (β) $d^{s-r} M_r > M_s$, we have $A > M_s$ for $\epsilon = 1$. Hence there exists a positive value $\epsilon_0 < 1$ for which $A = M_s$. Thus the r-th moment of Y_{ϵ_0} is equal to M_r and the s-th moment of Y_{ϵ_0} is equal to M_s. We have

$$P(Y_{\epsilon_0} < d) = p(0) + p(\sqrt[r]{M_r}) = \epsilon - \epsilon \frac{M_r}{t^r} + 1 - \epsilon = 1 - \epsilon \frac{M_r}{t^r} < 1.$$

Hence

$$a(d, t) < 1. \tag{7}$$

On account of (5) and (7) it follows from proposition III, that there exists a non-negative arithmetic random variable X belonging to the set $\Omega(t)$ such that $P(X < d) = a(d, t)$ and there exists at most one positive value $\delta(\neq d, \neq t)$ with positive probability. Hence $a(d, t)$ is equal to the *limes inferior* of the probabilities $P(Y < d)$ formed for all non-negative arithmetic random variables Y which have the following two properties:

(1) The r-th moment about the origin is equal to M_r and the s-th moment about the origin is equal to M_s

(2) There exists at most a single positive value $\delta(\neq d, \neq t)$ with positive probability.

Denote by Z a non-negative t-limited random variable with the properties (1), (2), and for which $P(Z < d) = a(d, t)$. The following equations hold

$$\begin{aligned} p(0) + p(\delta) + p(d) + p(t) &= 1 \\ p(\delta)\delta^r + p(d)d^r + p(t)t^r &= M_r \\ p(\delta)\delta^s + p(d)d^s + p(t)t^s &= M_s \end{aligned} \tag{8}$$

where $p(u)$ denotes the probability that $Z = u$.

From the last two equations of (8), we get

$$p(\delta) = \frac{M_r d^{s-r} - M_s + p(t)\,[t^s - t^r d^{s-r}]}{\delta^r(d^{s-r} - \delta^{s-r})} \tag{9}$$

$$p(d) = \frac{M_s - \delta^{s-r} M_r + p(t)\,[t^r \delta^{s-r} - t^s]}{d^r(d^{s-r} - \delta^{s-r})}. \tag{10}$$

Since $\frac{M_r}{d^r} > \frac{M_s}{d^s}$ and $t > d$, the numerator in (9) is positive. Since $0 \leq p(\delta) \leq 1$, the inequality

$$0 < \delta < d \tag{11}$$

must hold. Hence

$$p(\delta) > 0. \tag{12}$$

We shall show that $p(t) = 0$ if t is sufficiently large. For this purpose let us make the assumption $p(t) > 0$. We define a new random variable Z' as follows:

$$p'(t) = p(t) - \epsilon \text{ where } 0 < \epsilon < p(t)$$

$$p'(d) = p(d) - \epsilon \frac{t^r\delta^{s-r} - t^s}{d^r(d^{s-r} - \delta^{s-r})}$$

$$p'(\delta) = p(\delta) - \frac{\epsilon(t^s - t^r d^{s-r})}{\delta^r(d^{s-r} - \delta^{s-r})}$$

$$p'(0) = 1 - p'(\delta) - p'(d)\_p'(t)$$

and

$$p'(z) = 0 \text{ for all values } z \neq 0, \neq\delta, \neq d, \neq t.$$

$p'(u)$ denotes the probability that $Z' = u$.

The equations (8) remain satisfied if we substitute $p'(0)$, $p'(\delta)$, $p'(d)$, and $p'(t)$ for $p(0)$, $p(\delta)$, $p(d)$, and $p(t)$ respectively. Hence the r-th moment of Z' is equal to M_r and the s-th moment is equal to M_s. We have to show that Z' is in fact a random variable, that is to say, that the defined probabilities are ≥ 0 and ≤ 1. It is sufficient to show that the defined probabilities are non-negative, because the sum of them is equal to 1 and therefore they must be ≤ 1.

Obviously $p'(t)$ is >0. Since $t > d$ and according to (11) $d > \delta$, we have $p'(d) > p(d) > 0$. According to (12), $p(\delta)$ is positive. Hence for ϵ sufficiently small $p'(\delta)$ is also positive. We have to show that also $p'(0) \geq 0$. $p'(0)$ is given by

$$\begin{aligned} p'(0) &= 1 - p'(\delta) - p'(d) - p'(t) \\ &= 1 - p(\delta) - p(d) - p(t) + \epsilon\left[1 + \frac{t^r\delta^{s-r} - t^s}{d^r(d^{s-r} - \delta^{s-r})} + \frac{t^s - t^r d^{s-r}}{\delta^r(d^{s-r} - \delta^{s-r})}\right] \\ &= p(0) + \epsilon\frac{d^r\delta^r(d^{s-r} - \delta^{s-r}) + t^s(d^r - \delta^r) - t^r(d^s - \delta^s)}{d^r\delta^r(d^{s-r} - \delta^{s-r})}. \end{aligned}$$

Since $p(0) \geq 0$, $\epsilon > 0$, $d > \delta$ and $s > r$, this last expression is positive if t is sufficiently large. We may assume t so great that $p'(0) \geq 0$, because we want to calculate only

$$a_d = \lim_{t=\infty} a(d, t).$$

Now we shall show that

$$p'(d) + p'(t) > p(d) + p(t).$$

In fact

$$\begin{aligned} p'(d) + p'(t) - p(d) - p(t) &= \epsilon\left[\frac{t^s - t^r\delta^{s-r}}{d^r(d^{s-r} - \delta^{s-r})} - 1\right] \\ &= \epsilon\left[\frac{t^r}{d^r}\frac{t^{s-r} - \delta^{s-r}}{d^{s-r} - \delta^{s-r}} - 1\right] > 0. \end{aligned}$$

Then

$$p'(0) + p'(\delta) < p(0) + p(\delta) = a(d, t)$$

must follow. Since $p'(0) + p'(\delta) = P(Z' < d)$, we have a contradiction and therefore the assumption $p(t) > 0$ is reduced to an absurdity. Hence $p(t)$ must be equal to zero and $a(d, t) = a_d$. If we substitute zero for $p(t)$ in (8), (9), and (10) we obtain:

$$\tag{13} \begin{cases} p(0) + p(\delta) + p(d) = 1 \\ p(\delta)\delta^r + p(d)d^r = M_r \\ p(\delta)\delta^s + p(d)d^s = M_s \end{cases}$$

$$\tag{14} p(\delta) = \frac{M_r d^{s-r} - M_s}{\delta^r(d^{s-r} - \delta^{s-r})}$$

$$\tag{15} p(d) = \frac{M_s - M_r\delta^{s-r}}{d^r(d^{s-r} - \delta^{s-r})}.$$

We shall prove that $p(0) = 0$. For this purpose let us make the assumption $p(0) > 0$. Denote by δ_1 a positive number $<\delta$ and let us consider the arithmetic random variable Z' of degree 3 defined as follows:

$$p'(\delta_1) = \frac{M_r d^{s-r} - M_s}{\delta_1^r(d^{s-r} - \delta_1^{s-r})}$$

$$p'(d) = \frac{M_s - M_r\delta_1^{s-r}}{d^r(d^{s-r} - \delta_1^{s-r})}$$

$$p'(0) = 1 - p'(\delta_1) - p'(d).$$

The r-th moment of Z' is evidently equal to M_r and the s-th moment to M_s. Since $p(\delta) > 0$ according to (12), and $p(0) > 0$ by hypothesis, $p'(0)$ and $p'(\delta_1)$ will be greater than zero if δ_1 is sufficiently near to δ. The derivative of $p'(d)$ with respect to δ_1 is given by

$$\frac{1}{d^r}\,\frac{-M_r(s-r)\delta_1^{s-r-1}(d^{s-r} - \delta_1^{s-r}) + (s-r)\delta_1^{s-r-1}(M_s - M_r\delta_1^{s-r})}{(d^{s-r} - \delta_1^{s-r})^2}$$

$$= \frac{(s-r)\delta_1^{s-r-1}}{d^r(d^{s-r} - \delta_1^{s-r})^2}\,(M_s - M_r d^{s-r}).$$

Since $\dfrac{M_r}{d^r} > \dfrac{M_s}{d^s}$, the above expression is negative. Hence $p'(d)$ decreases with increasing δ_1. Since $\delta_1 < \delta$, we have

$$p'(d) > p(d) \geq 0$$

and therefore

$$1 - p'(d) < 1 - p(d) = a_d .$$

Since $1 - p'(d) = P(Z' < d)$, we have a contradiction and the assumption $p(0) > 0$ is proved an absurdity. Hence $p(0) = 0$, and $p(\delta) + p(d) = 1$. From (13), (14) and (15) we have

$$q(\delta) + p(d) = \frac{M_r d^s - M_s d^r + M_s \delta^r - M_r \delta^s}{\delta^r d^r (d^{s-r} - \delta^{s-r})} = 1 .$$

Hence

$$M_r d^s - M_s d^r + \delta^r (M_s - d^s) + \delta^s (d^r - M_r) = 0. \tag{16}$$

The equation (16) in δ has at most two positive roots, because the derivative of the left hand side of (16)

$$r\delta^{r-1}(M_s - d^s) + s\delta^{s-1}(d^r - M_r)$$

has exactly one positive root in δ. Since $\delta = d$ is a root of (16), the value of δ which we are seeking must be the second positive root of (16), which we shall denote by δ_0.

It can be easily shown that $\delta_0 \leq \sqrt[r]{M_r} < d$. In fact, for $\delta = 0$ the left hand side of (16) is positive on account of the assumption (β) and for $\delta = \sqrt[r]{M_r}$ it becomes equal to

$$M_s(M_r - d^r) - M_r^{\frac{s}{r}}(M_r - d^r) = \left(M_s - M_r^{\frac{s}{r}}\right)(M_r - d^r)$$

Since $M_s \geq M_r^{\frac{s}{r}}$ and recalling from (3) that $M_r < d^r$, the above expression is less than or equal to 0. Hence δ_0 lies between 0 and $\sqrt[r]{M_r} < d$.

Hence a_d is given by the expression

$$a_d = 1 - p(d) = 1 - \frac{M_s - M_r \delta_0^{s-r}}{d^r(d^{s-r} - \delta_0^{s-r})} . \tag{17}$$

For $s = 2r$ the root δ_0 can be easily calculated. We get

$$\delta_0 = \sqrt[r]{\frac{M_{2r} - d^r M_r}{M_r - d^r}} \tag{18}$$

If we substitute in (17) $2r$ for s and the right hand side of (18) for δ_0, then we get

$$\begin{aligned} a_d &= 1 - \frac{M_{2r} - M_r\left(\dfrac{M_{2r} - d^r M_r}{M_r - d^r}\right)}{d^r\left(d^r - \dfrac{M_{2r} - d^r M_r}{M_r - d^r}\right)} \\ &= 1 - \frac{(M_r - d^r)M_{2r} - M_r(M_{2r} - d^r M_r)}{d^r[d^r(M_r - d^r) - M_{2r} + M_r d^r]} \\ &= 1 - \frac{d^r(M_r^2 - M_{2r})}{d^r[2M_r d^r - d^{2r} - M_{2r}]} \\ &= 1 - \frac{M_r^2 - M_{2r}}{2M_r d^r - d^{2r} - M_{2r}} . \end{aligned}$$

Let us denote the non-negative number $M_{2r} - M_r^2$ by σ^2, then we obtain[7]

$$a_d = 1 - \frac{\sigma^2}{(d^r - M_r)^2 + \sigma^2}. \qquad (\sigma^2 = M_{2r} - M_r^2). \tag{19}$$

Let us compare the "sharp" limit given by (19) with the limit given by (2). If we substitute, in (2), $2r$ for s and d for $\xi \sqrt[r]{M_r}$ we have

$$b_d = 1 - \frac{M_{2r}}{d^{2r}} = 1 - \left(\frac{M_r}{d^r}\right)^2 - \frac{\sigma^2}{d^{2r}}$$

as a lower limit of the probability $P(-d < X < x_0 < d)$. We see that for small values of σ^2, b_d is considerably smaller than a_d.

Our results may be summarized in the following

THEOREM: *Denote by M_r the r-th and by M_s the s-th absolute moment of a random variable X about the point x_0, where $r < s$. For any positive value d denote by $P(-d < X < x_0 < d)$ the probability that $|X - x_0| < d$. The "sharp" lower limit a_d of $P(-d < X - x_0 < d)$ is defined as the limes inferior of the probabilities $P(-d < Y - x_0 < d)$ formed for all random variables Y for which the r-th moment about x_0 is equal to M_r and the s-th moment about x_0 is equal to M_s. We have to distinguish two cases.*

I. $\frac{M_r}{d^r} \leq \frac{M_s}{d^s}$. *In this case* $a_d = 1 - \frac{M_r}{d^r}$ *if* $\frac{M_r}{d^r} \leq 1$, *and* $a_d = 0$ *if* $\frac{M_r}{d^r} > 1$.

II. $\frac{M_r}{d^r} > \frac{M_s}{d^s}$. In this case a_d is given by

$$a_d = 1 - \frac{M_s - M_r\delta_0^{s-r}}{d^r(d^{s-r} - \delta_0^{s-r})}, \tag{17}$$

where δ_0 is the positive root $\neq d$ of the equation[8] in δ

$$M_r d^s - M_s d^r + \delta^r(M_s - d^s) + \delta^s(d^r - M_r) = 0.$$

For $s = 2r$ we have

$$\delta_0 = \sqrt[r]{\frac{M_{2r} - d^r M_r}{M_r - d^r}}.$$

If we substitute in (17) $2r$ for s and the above expression for δ_0, we obtain

$$a_d = 1 - \frac{\sigma^2}{(d^r - M_r)^2 + \sigma^2},$$

where $\sigma^2 = M_{2r} - M_r^2$.

COWLES COMMISSION, COLORADO SPRINGS.

[7] The case $s = 2r$ has been treated also by Cantelli. He demonstrated the formula (19) in quite another way, which cannot be generalized for the case $s \neq 2r$. Cantelli's formula and its demonstration are given in the book of M. Frechet, *Generalities sur Probabilities. Variables Aleatoires*, Paris, 1937, pp. 123–126.

[8] As has been shown, there exists exactly one positive root $\neq d$ of the equation considered.

A MODIFICATION OF BAYES' PROBLEM

By R. v. Mises

The classical Bayes problem can be stated as follows. We consider an urn which contains white and black balls (or balls designated by 0 and 1). The probability p for drawing a black ball is unknown. But there is given a probability function $F(x)$ representing the *a priori* probability for the inequality $p \leqq x$. We draw n times from the urn (returning each time the extracted ball) and get a black ball m times and a white one $n - m$ times. Now, after this experiment, we ask for the *a posteriori* probability $P_n(x)$ for the relation $p \leqq x$.

The solution proposed by Bayes can be written in a slightly generalized form:

$$P_n(x) = K \int_0^x p^m(1 - p)^{n-m} \, dF(p) \tag{1}$$

where K is a constant to be found by means of the condition

$$P_n(1) = 1. \tag{1'}$$

We are interested in the behaviour of $P_n(x)$ if n tends to ∞ under the condition

$$\lim_{n \to \infty} \frac{m}{n} = \alpha. \tag{2}$$

Laplace found in the case of a priori equipartition $F(x) = x$, and I proved in 1919[1] for any derivable $F(x)$, that $P_n(x)$ tends to a normal distribution:

$$\lim_{n \to \infty} \left[P_n(x) - \frac{1}{\sqrt{\pi}} \int_{-\infty}^u e^{-u^2} \, du \right] = 0 \tag{3}$$

with

$$u = H_n(x - A_n)$$

$$A_n = \alpha, \qquad \frac{1}{2H_n^2} = \frac{\alpha(1 - \alpha)}{n}. \tag{4}$$

It is easily seen from (3) and (4) that

$$\lim_{n \to \infty} P_n(x) = \begin{cases} 0 & \text{if } x < \alpha \\ 1 & \text{“ } x > \alpha. \end{cases} \tag{5}$$

Let us now consider a slightly modified form of the problem.[2] Instead of one

---

[1] *Mathematische Zeitschrift*, vol. 4 (1919) p. 92. Cf. my textbook *Wahrscheinlichkeitsrechnung und ihre Anwendungen*, Wien-Leipzig 1931, p. 158. Later I proved the Laplace-Bayes theorem for a more general class of $F(x)$: *Monatshefte für Mathematik und Physik*, vol. 43 (1936) pp. 105–128.

[2] This modified problem has been treated by S. Bochner, *Annals of Math.*, Vol. 37, 1936, p. 816.

urn we suppose there are given n urns each containing white and black balls. The probability p_ν for drawing a black ball from the ν^{th} urn is unknown, but is subject to an a priori probability function $F(x)$ which furnishes the a priori probability for the relation $p_\nu \leqq x$, independently of ν. We assume that on drawing one ball from every urn a black ball appears m times and a white ball $n - m$ times. Putting

$$\frac{p_1 + p_2 + \cdots + p_n}{n} = p, \tag{6}$$

we ask for the a posteriori probability $P_n(x)$ for the relation $p \leqq x$.

The Bayes formula (1) must now be replaced by

$$\begin{aligned} P_n(x) = K' \int\int_{p_1+p_2+\cdots p_n \leqq nx} \cdots \int p_1 p_2 \cdots p_m (1 - p_{m+1})(1 - p_{m+2}) \\ \cdots (1 - p_n)\, dF(p_1) \cdots dF(p_n) \end{aligned} \tag{7}$$

where K' is a constant determined by (1′). It is very easy to examine the asymptotic character of (7). We shall prove the following

THEOREM: *If the first three moments of the a priori distribution $F(x)$*

$$b_\nu = \int_0^1 x^\nu\, dF(x), \qquad \nu = 1, 2, 3 \tag{8}$$

exist and if the dispersion $b_2 - b_1^2$ is different from 0, the a posteriori probability $P_n(x)$ tends for $n \to \infty$ under the condition (2) to the normal distribution (3) with

$$\begin{aligned} A_n &= \alpha \frac{b_2}{b_1} + (1 - \alpha) \frac{b_1 - b_2}{1 - b_1} \\ \frac{1}{2H_n^2} &= \frac{1}{n}\left[\alpha \frac{b_1 b_3 - b_2^2}{b_1^2} + (1 - \alpha) \frac{(b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2}{(1 - b_1)^2}\right]. \end{aligned} \tag{9}$$

In order to prove the theorem we write

$$\begin{aligned} V_\nu(p_\nu) &= \frac{1}{b_1} \int_0^{p_\nu} x\, dF(x), \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{1}{1 - b_1} \int_0^{p_\nu} (1 - x)\, dF(x), \quad \text{if } \nu = m + 1, m + 2, \cdots n. \end{aligned} \tag{10}$$

Then formula (7) becomes

$$P_n(x) = C \int\int_{p_1+p_2+\cdots p_n \leqq nx} \cdots \int dV_1(p_1)\, dV_2(p_2) \cdots dV_n(p_n). \tag{11}$$

Each $V_\nu(p_\nu)$ is a distribution function, i.e. a non-decreasing function with $V_\nu(-\infty) = 0$, $V_\nu(\infty) = 1$. Therefore the constant C in (11) is equal to 1 and

the integral represents the distribution function for the arithmetical mean $(p_1 + p_2 + \cdots p_n)/n$. According to the *Central Limit Theorem* of the theory of probability $P_n(x)$ will converge towards a normal distribution when certain conditions are satisfied. In every case, if a_ν, s_ν^2 denote the mean value and the dispersion associated with $V_\nu(x)$, then the mean value A_n and the dispersion S_n^2 associated with $P_n(x)$ will be defined by

$$A_n = \frac{1}{n}\sum_{\nu=1}^{n} a_\nu, \qquad S_n^2 = \frac{1}{n^2}\sum_{\nu=1}^{n} s_\nu^2. \tag{12}$$

We find from (10)

$$\begin{aligned} a_\nu &= \int_0^1 x\,dV_\nu(x) = \frac{1}{b_1}\int_0^1 x^2\,dF(x) = \frac{b_2}{b_1}, \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{1}{1-b_1}\int_0^1 x(1-x)\,dF = \frac{b_1 - b_2}{1 - b_1}, \quad \text{if } \nu = m+1, \cdots n \end{aligned} \tag{13}$$

$$\begin{aligned} s_\nu^2 &= \int_0^1 x^2\,dV_\nu(x) - a_\nu^2 = \frac{b_3}{b_1} - \frac{b_2^2}{b_1^2}, \quad \text{if } \nu = 1, 2, \cdots m \\ &= \frac{b_2 - b_3}{1 - b_1} - \frac{(b_1 - b_2)^2}{(1-b_1)^2}, \quad \text{if } \nu = m+1, \cdots n. \end{aligned} \tag{14}$$

We supposed the dispersion of $F(x)$ to be different from zero. It follows that

$$b_1 \neq 0,\ 1 - b_1 \neq 0,\ b_3 b_1 - b_2^2 \neq 0,\ (b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2 \neq 0. \tag{15}$$

For $b_1 = 0$ would imply that $dF(x) = 0$ for all $x > 0$ and $1 - b_1 = 0$ that $dF(x) = 0$ for all $x < 1$; in both cases the dispersion would be zero. On the other hand, it is easily seen that the relation $b_3b_1 - b_2^2 = 0$ is not compatible with the condition of a non-vanishing a priori dispersion and that the same is true for the relation $(b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2 = 0$.

The total dispersion Σs_ν^2 is equal to the sum of m times the value $(b_3b_1 - b_2^2)/b_1^2$ and $n - m$ times the value $[(b_2 - b_3)(1 - b_1) - (b_1 - b_2)^2]/(1 - b_1)^2$.

Thus we see that under the condition (2) the sum Σs_ν^2 tends to ∞, while the ratio $s_\nu^2/\Sigma s_\nu^2$ tends to zero, if n increases infinitely. These are sufficient conditions for the validity of the Central Limit Theorem.[3] The values given for A_n and H_n^2 in (9) follow from (12), (13), (14) and the well known relation $2H_n^2 S_n^2 = 1$.

S. Bochner in his previously quoted paper found, in a more complicated manner, the value of A_n and only showed that $P_n(x)$ tends to zero if $x < A_n$ and to 1 if $x > A_n$.

Examples. If we assume the a priori probability to be uniform, i.e. $F(x) = x$, we have

$$b_1 = \tfrac{1}{2}, \qquad b_2 = \tfrac{1}{3}, \qquad b_3 = \tfrac{1}{4}$$

and therefore from (9)

[3] Cf. H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tract in Mathematics and Mathematical Physics, No. 36, 1937, p. 56.

$$A_n = \tfrac{1}{3}(\alpha + 1), \qquad \frac{1}{2H_n^2} = \frac{1}{18n}.$$

A more general case is that of a more concentrated a priori probability function

$$F'(x) = Cx^k(1 - x)^l, \qquad C = \frac{(k + l + 1)!}{k!\,l!}.$$

Here we find

$$b_1 = \frac{k + 1}{k + l + 2}, \qquad b_2 = \frac{(k + 1)(k + 2)}{(k + l + 2)(k + l + 3)},$$

$$b_3 = \frac{(k + 1)(k + 2)(k + 3)}{(k + l + 2)(k + l + 3)(k + l + 4)}$$

and the values of A_n and H_n^2 are

$$A_n = \frac{\alpha + k + 1}{k + l + 3}, \qquad \frac{1}{2H_n^2} = \frac{\alpha(l - k) + (k + 1)(l + 2)}{n(k + l + 3)^2(k + l + 4)}.$$

By introducing the moments of $F(x)$ relative to the mean value, i.e.

$$\begin{aligned} B_2 &= \int_0^1 (x - b_1)^2\,dF = b_2 - b_1^2, \\ B_3 &= \int_0^1 (x - b_1)^3\,dF = b_3 - 3b_1b_2 + 2b_1^3 \end{aligned} \tag{16}$$

we can transform the general formulas (9) into

$$\begin{aligned} A_n &= b_1 + \frac{B_2}{b_1(1 - b_1)}(\alpha - b_1) \\ \frac{1}{2H_n^2} &= \frac{1}{n}\left[B_2 + B_3\,\frac{\alpha - b_1}{b_1(1 - b_1)} - B_2^2\,\frac{b_1^2 + \alpha(1 - 2b_1)}{b_1^2(1 - b_1)^2}\right]. \end{aligned} \tag{17}$$

The first of these equations shows that the a posteriori mean value A_n (for all n) is equal to the a priori mean value b_1, if the experimental mean m/n or α coincides with the latter. On the other hand, in the case of a symmetric a priori distribution ($b_1 = \frac{1}{2}$, $B_3 = 0$) the second equation is reduced to

$$\frac{1}{2H_n^2} = \frac{1}{n}(B_2 - 4B_2^2).$$

On the whole it is remarkable that the influence of the a priori probability does not vanish for $n \to \infty$, in the case of our modified Bayes problem.[4] The explanation of this fact is to be found in a more generalized theory of the inverse problems in probability.

UNIVERSITY OF ISTANBUL, TURKEY.

[4] Cf. my papers quoted in footnote 1.

ON THE PROBABILITY THEORY OF ARBITRARILY LINKED EVENTS

By Hilda Geiringer

1. **Introduction.** The classical Poisson problem can be stated as follows: Let $p_1, p_2, \cdots p_n$ be the probabilities of n *independent* events $E_1, E_2, \cdots E_n$ respectively; i.e. the probability of the simultaneous occurrence of E_i and E_j is equal to $p_i p_j$, that of E_i, E_j, E_k is equal to $p_i p_j p_k$ and so on. We seek the probability $P_n(x)$ that x of the events shall occur. If, $p_1 = p_2 = \cdots = p_n$ the problem is known as the Bernoulli problem.

More generally the n events may be regarded as *dependent*. Let p_{ij} be the probability of the simultaneous occurrence of E_i and E_j; p_{ijk} that of E_i, E_j, E_k and finally $p_{12\cdots n}$ that of $E_1, E_2, \cdots E_n$. There shall arise again the problem of determining the probability $P_n(x)$ that x of the n events will take place.[1] Furthermore the asymptotic behaviour of $P_n(x)$ for large n can be studied; and we shall especially be interested in the problem of the convergence of $P_n(x)$ towards a normal distribution or a Poisson distribution.

Even in the general case which we just explained, the sums

$$S_1 = \sum_{i=1}^{n} p_i, \qquad S_2 = \sum_{i,j=1}^{n} p_{ij}, \cdots S_n = p_{12\cdots n}$$

of our probabilities differ only by constant factors from the *factorial moments* $M_n^{(1)}, M_n^{(2)}, \cdots M_n^{(n)}$ of $P_n(x)$. For we have

$$S_\nu = \frac{1}{\nu!} M_n^{(\nu)} = \frac{1}{\nu!} \sum_{x=\nu}^{n} x(x-1) \cdots (x-\nu+1) P_n(x).$$

Starting from this remark the author has, in earlier papers, [8, 9, 10] established a theory of the asymptotic behaviour of $P_n(x)$, making use of the theory of moments. The criterion for the convergence of $P_n(x)$ towards the normal—or the Poisson—distribution consists of certain conditions[2] which the S_ν must satisfy.

In the following section a concise statement of the whole problem will be given, independently of the author's earlier publications. For the convergence towards the normal distribution we shall be able to establish a theorem under wider conditions in a manner which seems to be simpler. Finally, some applications of the theory will be considered.

---

[1] See, for instance, references [1]-[7] at end of paper.

[2] Using the "theorem of the continuity of moments," Professor v. Mises [11] established sufficient conditions for the convergence of $P_n(x)$ towards a Poisson distribution in the case of the problem of "iterations." However, his reasoning can be applied to the general case without much difficulty.

2. **Formulation of the problem.** Let us consider the n-dimensional *collective* (Kollectiv) consisting of a sequence of any n trials. In the simplest case these trials will be *alternatives*, i.e. for every trial there will exist only two results, which we may denote by "occurrence," "non-occurrence" or by "1," "0." The single trial may eventually be composed in various manners. For instance we may draw $m > n$ times from an urn, which contains counters, bearing in arbitrary proportions numbers from 0 to 9. The first "event" E_1 may consist of the fact that the first three extracted counters bear even numbers; the second trial E_2 will be regarded as successful, if the sum of the counters extracted at the second, third and fourth drawings is greater than five, etc. In every case the result of the n trials will be expressed by n numbers, each of them equal to 0 or 1. The result (1, 1, 0, 0, 0, $\cdots$ 1), for instance, means that the first, the second, and the last trial were successful, the third, fourth, $\cdots$ unsuccessful, and we have an arithmetical probability distribution $v(x_1, x_2, \cdots x_n)$ $(x_k = 0, 1;$ $k = 1, 2, \cdots n)$, where

(1) $$\sum_{x_1} \cdots \sum_{x_n} v(x_1, x_2, \cdots x_n) = 1.$$

Instead of the $2^n - 1$ values of v we will deal with certain groups of *partial sums* of them; the first is

$$\sum \cdots \sum v(x_1, x_2, \cdots x_{i-1}, 1, x_{i+1} \cdots x_n) = p_i \qquad (i = 1, 2, \cdots n)$$

where p_i is the probability that the i-th trial will be successful. In an analogous manner let p_{ij} be the probability that the i-th and the j-th trial are both sucessful, p_{ijk} the probability that the i-th, j-th and k-th trials are simultaneously successful. Let us provisionally denote by $\Sigma^{(i)}$ an $(n - 1)$-tuple sum over all variables, except x_i, by $\Sigma^{(i,j)}$ an $(n - 2)$-tuple sum over all variables except x_i and x_j etc. We shall then have:

(2)
$$\begin{aligned}
p_i &= \sum{}^{(i)} v(x_1, \cdots x_{i-1}, 1, x_{i+1}, \cdots x_n) \\
p_{ij} &= \sum{}^{(i,j)} v(x_1, \cdots x_{i-1}, 1, x_{i+1}, \cdots x_{j-1}, 1, x_{j+1}, \cdots x_n) \\
&\cdots\cdots\cdots\cdots\cdots \\
p_{12\cdots n} &= v(1, 1, \cdots 1).
\end{aligned}$$

In the following these probabilities p_i, p_{ij}, p_{ijk} $\cdots$ will be assumed as *directly given*. There are

$$\binom{n}{1} + \binom{n}{2} + \cdots \binom{n}{n-1} + \binom{n}{n} = 2^n - 1$$

values of this kind and it is easily seen, that the partial sums (2) are *linearly independent*.

If, especially, the probability $v(x_1, x_2, \cdots x_n)$ *depends only on the number* of zeros amongst $x_1, x_2, \cdots x_n$, i.e. if

$$v(1, 0, \cdots 0) = v(0, 1, 0, \cdots 0) = \cdots = v(0, 0, \cdots 1)$$
$$v(1, 1, \cdots 0) = v(1, 0, 1, \cdots 0) = \cdots = v(0, 0, \cdots 0, 1, 1)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$

the value of p_i is independent of i, the value of p_{ij} independent of i and j, and so on:

$$p_1 = p_2 = \cdots = p_n$$
$$p_{12} = p_{23} = \cdots = p_{n-1,n}$$

In the particular case of *independent* events we have only to deal with n probabilities, namely $p_1, p_2, \cdots p_n$. We have indeed $p_{ij} = p_i p_j$; $p_{ijk} = p_i p_j p_k \cdots p_{12\cdots n} = p_1 p_2 \cdots p_n$.

In the case of *chains* however, we need only know $(2n - 1)$ values, namely $p_1, p_2, \cdots p_n$; $p_{12}, p_{23}, \cdots p_{n-1,n}$. The other p_{ij}, and the $p_{ijk}, \cdots p_{12\cdots n}$ can be expressed in terms of the above probabilities.

Returning now to the general case it is easily seen that in the expression for $P_n(x)$ the $p_i, p_{ij} \cdots$ will appear only in the following combinations

$$S_n(0) = 1, \qquad S_n(1) = \sum_i^{1\cdots n} p_i, \qquad S_n(2) = \sum_{i,j}^{1\cdots n} p_{ij}, \cdots S_n(n) = p_{12\cdots n}. \tag{3}$$

Indeed, at the basis of the solution of the "problem of sums," there are the following relations [11] between the $S_n(z)$ and the $P_n(x)$.

$$S_n(z) = \sum_{x=z}^{n} \binom{x}{z} P_n(x) \qquad \begin{pmatrix} x = 0, \cdots n \\ z = 0, \cdots n \end{pmatrix} \tag{4}$$

The linear equations (4) may be solved (by recurrence) for the $P_n(x)$ and we find the important result that

$$P_n(x) = \sum_{z=x}^{n} (-1)^{z+x} \binom{z}{x} S_n(z) \tag{5}$$

Let $M_n^{(z)}$ be the z-th *factorial moment* of $P_n(x)$, i.e.

$$M_n^{(z)} = \sum_{x=z}^{n} x(x-1) \cdots (x - z + 1) P_n(x). \tag{6}$$

Making use of (4) and (6) we obtain

$$M_n^{(z)} = z! S_n(z). \tag{7}$$

Our aim is to obtain information concerning the asymptotic behaviour of $P_n(x)$ by studying that of the moments of $P_n(x)$. The moments however are easily seen to be given in terms of the $S_n(z)$.

3. **The asymptotic behavior of $P_n(x)$. Convergence towards the normal distribution.**

a. THE PRINCIPAL THEOREM. According as the mean value

$$M_n^{(1)} = S_n(1) = a_n = \sum_{x=1}^{n} xP_n(x) \tag{8}$$

remains bounded or not for indefinitely increasing n, there are two types of passage to a limit. In the first case the distribution will converge (under certain conditions) towards a *Poisson* distribution; in the second case it will approach (under certain conditions) a normal distribution. As regards the convergence towards the *Poisson* distribution the author has published [9] a sufficient condition which seems to be quite simple and general. We shall, however, not resume this problem in the present paper.

We propose, indeed, to prove in the following pages a new theorem concerning the convergence of

$$V_n(x) = \sum_{t \leq x} P_n(t)$$

towards a *normal distribution*.

For this purpose we introduce the following function of the discontinuous variable $z = 0, 1, 2, \cdots n$

$$g_n(z) = \frac{z+1}{a_n}\frac{S_n(z+1)}{S_n(z)} \tag{9}$$

or, more concisely written $g_z = \dfrac{z+1}{a}\dfrac{S_{z+1}}{S_z}$, where $S_n(z)$ is defined by (3). Putting $z = a_n u$, let us consider

$$g_n(a_n u) = h_n(u) \tag{10}$$

where u is regarded as a *continuous* variable in the interval from 0 to ϵ. ($\epsilon > 0$.)

Denoting the variance of x for $V_n(x)$ by $M_2 = s_n^2$ we shall prove the

THEOREM: *Let the function $h_n(u)$, defined by (10) satisfy the following conditions:*

(i) *If n is sufficiently large, $h_n(u)$ admits derivatives of every order in the interval* $(0, \epsilon)$

(ii) *At $u = 0$, the first derivative of $h_n(u)$ has a limit, for $n \to \infty$, which is different from* -1.

(iii) *If u is in the interval $(0, \epsilon)$ the k-th derivative of $h_n(u)$ remains, for every k, inferior to a bound N_k which is independent of n.*

Then

$$\lim_{n\to\infty} V_n(a_n + ys_n\sqrt{2}) = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{y} e^{-x^2}dx \tag{11}$$

We shall see that in many applications these conditions may reasonably be assumed as satisfied.

b. DEMONSTRATION OF THE THEOREM.

In order to prove the principal theorem, stated above, we shall at first deduce some properties of the (finite) *differences* of $g_n(z)$ ($z = 0, 1, \cdots$) from the assumptions (i), (ii), (iii) which deal with the *derivatives* of $h_n(u)$. Indeed, the κ-th

difference of $g_n(z)$ with respect to z, (which contains the values of $g_n(z)$ for $z = 0, 1, \cdots \kappa$), differs only by the factor a_n^κ from the κ-th divided difference of $h_n(u)$, with respect to u (which is formed by the values of $h_n(u)$ for $u = 0$, $\frac{1}{\alpha_n}, \cdots \frac{\kappa}{\alpha_n}$. Let $n > \kappa$ and so large that $\kappa/a_n < \epsilon$; then all u-values used in the formation of the κ-th divided difference of $h_n(u)$ will be in the interval $(0, \epsilon)$. Now, as it is well known, the absolute value of any divided difference of order κ can not be larger than the largest derivative in an interval which contains all the abscissae, used in the formation of the divided difference. But according to hypothesis (iii) the κ-th derivatives of $h_n(u)$ in $(0, \epsilon)$ are all inferior to N_κ. Therefore[3] we have

$$| a_n^\kappa \Delta^\kappa g_n(z) | < N_\kappa \tag{12}$$

and for every $\gamma > 0$

$$\lim_{n\to\infty} \alpha_n^{\kappa-\gamma} \mathop{\Delta^\kappa}_{z=0} g_n(z) = 0. \tag{13}$$

On the other hand from condition (ii) it follows, as is easily seen, that

$$\lim_{n\to\infty} \alpha_n \mathop{\Delta}_{z=0} g_n(z) = a_n[g_n(1) - g_n(0)] = c \neq -1. \tag{14}$$

The equations (13) and (14) imply but *finite differences of* $g_n(z)$.

Let us now introduce certain new moments F_ν which we could call "factorial moments about the mean." They are indeed related to the factorial moments $M^{(\nu)}$ in exactly the same way as the moments M_ν about the mean are related to the moments M_ν^0 about the origin. Writing, S_z, a and g_z instead of $S_n(z)$, a_n and $g_n(z)$, we set

$$\begin{aligned} F_\nu &= \mathop{\Delta^\nu}_{z=0} (M^{(\nu)} a^{\nu-z}) = M^{(\nu)} - \nu M^{(\nu-1)}a + \binom{\nu}{2} M^{(\nu-2)}a^2 - \cdots \pm a^\nu \\ &= \nu! S_\nu - \nu! S_{\nu-1} a + \binom{\nu}{2}(\nu-2)! S_{\nu-2} a^2 - \cdots \pm a^\nu \end{aligned} \tag{15}$$

where, particularly,

$$F_0 = 1, \qquad F_1 = 0. \tag{16}$$

From (15) we have:

$$\begin{aligned} M^{(\nu)} &= \nu! S_\nu = \sum_{z=0}^{\nu} F_{\nu-z} \binom{\nu}{z} a^z \\ &= F_\nu + \nu F_{\nu-1} a + \binom{\nu}{2} F_{\nu-2} a^2 + \cdots + \binom{\nu}{\nu-2} F_2 a^{\nu-2} + a^\nu \end{aligned} \tag{17}$$

Let us begin by proving the following

[3] If we only want to deduce (13) it is sufficient to suppose that N_k (without being independent of n) increases more slowly than any power of a_n.

LEMMA I: *It follows from* (13) *and* (14) *that we have for the* F_ν *defined by* (15)

$$\lim_{n\to\infty} \frac{F_\nu}{a^{\frac{1}{2}\nu}} = G_\nu = \begin{cases} 0 & \text{if } \nu \text{ odd} \\ 1\cdot 3 \cdots (\nu-1)c^{\frac{1}{2}\nu} & \text{if } \nu \text{ even.} \end{cases} \tag{18}$$

First we conclude from (15) and (14) that (18) is true for $\nu = 1$ and $\nu = 2$. In order to prove (18) for *every* ν, we shall point out, that

$$\lim_{n\to\infty} \frac{F_\nu}{a^{\frac{1}{2}\nu}} = (\nu - 1)c \lim_{n\to\infty} \frac{F_{\nu-2}}{a^{\frac{1}{2}(\nu-2)}} \cdots \qquad (\nu = 2, 3, \cdots) \tag{19}$$

Setting

$$f_z = g_z - 1 \quad \text{and} \quad m_z = \frac{S_z z!}{a^z} = \frac{M^{(z)}}{a^z} \tag{20}$$

we get

$$g_z = \frac{m_{z+1}}{m_z} \tag{21}$$

and

$$\begin{aligned} \Delta m_z &= m_z f_z \\ \Delta^\nu m_z &= \Delta^{\nu-1}(m_z f_z) \end{aligned} \qquad (z = 0, 1, 2, \cdots) \tag{22}$$

But according to (15) we have

$$\underset{z=0}{\Delta^\nu} m_z = \frac{1}{a^\nu} F_\nu \tag{23}$$

and therefore

$$\frac{F_\nu}{a^{\frac{1}{2}\nu}} = a^{\frac{1}{2}\nu} \underset{z=0}{\Delta^{\nu-1}} (m_z f_z) = a^{\frac{1}{2}\nu} \sum f_{\alpha\beta} \Delta^\alpha m_z \Delta^\beta f_z = \sum f_{\alpha\beta} \frac{F_\alpha}{a^{\frac{1}{2}\alpha}} a^{\frac{1}{2}(\nu-\alpha)} \Delta^\beta f_z \tag{24}$$

$$(\alpha + \beta \geqq \nu - 1;\ \alpha \leqq \nu - 1,\ \beta \leqq \nu - 1).$$

Here we have made use of the fact that the κ-th difference of a product uv can be transformed in a finite sum $\Sigma S_{\alpha\beta}\Delta^\alpha u \Delta^\beta v$ where α and β are non-negative integers and $\alpha \leqq \kappa$, $\beta \leqq \kappa$. $\Big($If we concern ourselves with derivatives and not with finite differences, we have, $\alpha + \beta = \kappa$ and $S_{\alpha\beta} = \binom{\kappa}{\alpha}\Big)$. Suppose

$$\alpha + \beta > \nu - 1.$$

Then $\beta \geqq \nu - \alpha$; therefore, as $\nu > \alpha$ we have $\beta > \dfrac{\nu - \alpha}{2}$. Since $\Delta^\beta f_z = \Delta^\beta g_z$ the product $a^{\frac{1}{2}(\nu-\alpha)} \underset{z=0}{\Delta^\beta} f_z$ *converges toward* zero, in accordance with (13), whereas the factor $S_{\alpha\beta} \dfrac{F_\alpha}{a^{\frac{1}{2}\alpha}}$ remains bounded for every $\alpha < \nu$. Now suppose

$$\alpha + \beta = \nu - 1.$$

Then $\beta = \nu - 1 - \alpha$. First let $\alpha < \nu - 2$; then $\beta = \nu - 1 - \alpha > \frac{\nu - \alpha}{2}$. Thus $a^{\frac{1}{2}(\nu-\alpha)} \Delta^\beta_{z=0} f_z$ converges again towards zero, whereas the other factors are bounded as before. Next, if $\alpha = \nu - 1$, then $\beta = 0$ and $\Delta^0_{z=0} f_z = f_0 = 0$. Thus the corresponding term of our sum is equal to zero. Finally if $\alpha = \nu - 2$, then $\beta = 1$, and $S_{\alpha\beta} = \nu - 1$. The corresponding term of the sum (24) will be

$$(\nu - 1) \lim_{n\to\infty} \frac{F_{\nu-2}}{a^{\frac{1}{2}(\nu-2)}} \cdot \lim_{n\to\infty} a \mathop{\Delta}_{z=0} f_z = (\nu - 1)c \lim_{n\to\infty} \frac{F_{\nu-2}}{a^{\frac{1}{2}(\nu-2)}}$$

which completes the proof of Lemma I.

We shall now establish a relation between the *factorial moments* about the mean F_ν and the *ordinary moments* about the mean M_ν. To an expression of the form

$$ca^\rho F_\alpha \tag{25}$$

(where the constant c is independent of n) let us attribute a "weight" $\rho + \frac{\alpha}{2}$. Then we shall prove the following *lemma*

LEMMA II: *Let* $\nu = 2\mu$(ν even), $\nu = 2\mu + 1$ (ν odd) *and*

$$\alpha_\rho = \frac{\nu!}{(\nu - 2\rho)!\, 2^\rho \rho!}. \tag{26}$$

Then

$$M_\nu - \sum_{\rho=0}^{\mu} \alpha_\rho a^\rho F_{\nu-2\rho} \tag{27}$$

is equal to a finite sum of terms of the form (25), *each of which has a weight less than* $\nu/2$.

To prove this lemma we begin by expressing the M_ν in terms of the factorial moments $M^{(\rho)}$. We shall then express the $M^{(\rho)}$ by the F_z. Now, let $s_{\kappa z}$ be the "*Stirling* numbers of second kind," i.e., putting

$$x^{(z)} = x(x - 1) \cdots (x - z + 1) \tag{28}$$

we have

$$x^z = \sum_{\kappa=0}^{z} s_{\kappa z} x^{(z-\kappa)} \qquad (z = 0, 1, 2, \cdots) \tag{29}$$

Then by an elementary calculation we obtain

$$M_\nu = \sum_{\rho=0}^{\nu} M^{(\nu-\rho)} \left[s_{\rho\nu} - \nu a s_{\rho-1,\nu-1} + \binom{\nu}{2} a^2 s_{\rho-2,\nu-2} - \cdots \pm \binom{\nu}{\rho} a^\rho \right]. \tag{30}$$

If we now introduce the F_κ we get

$$
(31) \quad \begin{aligned} M_\nu = \sum_{\rho=0}^{\nu-1} \sum_{r=0}^{\nu-\rho} F_{\nu-r-\rho} a^r \Bigg[\binom{\nu-\rho}{r} s_{\rho\nu} - \binom{\nu}{1}\binom{\nu-\rho-1}{r-1} s_{\rho,\nu-1} \\ + \binom{\nu}{2}\binom{\nu-\rho-2}{r-2} s_{\rho,\nu-2} - \cdots \pm \binom{\nu}{r} s_{\rho,\nu-r} \Bigg]. \end{aligned}
$$

Furthermore we may easily verify that

$$
(32) \quad \begin{aligned} \binom{\nu-\rho-x}{r-x} = \binom{\nu-\rho}{r} - \binom{\nu-\rho-1}{r} x \\ + \frac{1}{2!}\binom{\nu-\rho-2}{r} x^{(2)} + \cdots \pm \frac{1}{(\nu-r-\rho)!} x^{(\nu-r-\rho)}. \end{aligned}
$$

But the $s_{\kappa z}$ for $z = 0, 1, 2, \cdots$ are equal to the values of a polynomial in z, of degree 2κ, the highest term of which is equal to $\dfrac{z^{2\kappa}}{\kappa!\,2^\kappa}$. The degree of the product

$$
(33) \quad \binom{\nu-\rho-x}{r-x} s_{\rho,\nu-x} = \varphi(x)
$$

is therefore equal to $(\nu - r - \rho) + 2\rho = \nu - r + \rho$. On the other hand the expression between brackets in the right hand member of (31) is nothing other than the ν-th difference of $\zeta(x)$. (The missing terms of this difference are indeed equal to zero, the corresponding $s_{\kappa\nu}$ being equal to zero.)

This ν-th difference will certainly vanish if

$$
\nu - r + \rho < \nu \text{ i.e. } r > \rho.
$$

Now, let $r = \rho$. Then the ν-th difference, i.e. the coefficient α_ρ of $F_{\nu-r-\rho}a^r = F_{\nu-2\rho}a^\rho$ in (31), is equal to $\nu!$ multiplied by the coefficient of $x^{\nu-2\rho}$ in $\varphi(x)$:

$$
\alpha_\rho = \nu! \frac{1}{(\nu-2\rho)!} \frac{1}{2^\rho \rho!}.
$$

Finally, let $r < \rho$. Then the weight of $F_{\nu-r-\rho}a^r$ is inferior to $\nu/2$. We have thus established Lemma II.

We have for instance for $\nu = 1, 2, 3, 4, 5$

$$
M_1 = F_1 = 0,\ M_2 = F_2 + a,\ M_3 = F_3 + 3F_2 + a
$$

$$
M_4 = (F_4 + 6\,aF_2 + 3\,a^2) + 6\,F_3 + (7F_2 + a)
$$

$$
M_5 = (F_5 + 10\,F_3 a) + (10\,F_4 + 40\,F_2 a + 10\,a^2) + 25\,F_3 + (15\,F_2 + a)
$$

Inversely in an analogous manner, we can express F_ν by the

$$
M_\rho (\rho = 1, 2, \cdots \nu).
$$

We can now terminate our demonstration by proving the following
Lemma III: *If the conditions* (18) *are satisfied, then*

(34) $$\lim_{n\to\infty} \frac{M_\nu}{M_2^{\frac{1}{2}\nu}} = H_\nu = \begin{cases} 0 \cdots \nu \text{ odd} \\ 1\cdot 3 \cdots (\nu - 1) \cdots \nu \text{ even.} \end{cases}$$

First the equation (18) for $\nu = 2$ gives

$$\lim_{n\to\infty} \frac{F_2}{a} = \lim_{n\to\infty} \frac{M_2 - a}{a} = c$$

thus

(35) $$\lim_{n\to\infty} \frac{M_2}{a} = 1 + c \qquad (c \neq -1).$$

It is therefore obviously sufficient to prove the relation

(36) $$\lim_{n\to\infty} \frac{M_\nu}{a^{\frac{1}{2}\nu}} = H_\nu(1 + c)^{\frac{1}{2}\nu}.$$

Putting $\nu = 2\mu$ and $\nu = 2\mu + 1$ respectively we obtain however from our lemma

$$\frac{M_\nu}{a^{\frac{1}{2}\nu}} = \sum_{\rho=0}^{\mu} \alpha_\rho a^{\rho - \frac{1}{2}\nu} F_{\nu-2\rho} + Ra^{-\frac{1}{2}\nu}.$$

Here R represents a finite sum of terms of the form (25), of "weight" inferior to $\frac{\nu}{2}$. But by virtue of (18) such a term, divided by $a^{\frac{1}{2}\nu}$ converges towards zero and we obtain

(37) $$\lim_{n\to\infty} \frac{M_\nu}{a^{\frac{1}{2}\nu}} = \sum_{\rho=0}^{\mu} \alpha_\rho \lim_{n\to\infty} \frac{F_{\nu-2\rho}}{a^{\frac{1}{2}\nu-\rho}} = \sum_{\rho=0}^{\mu} \frac{\nu!}{(\nu - 2\rho)!2^\rho \rho!} G_{\nu-2\rho}.$$

For an odd ν, $G_{\nu-2\rho}$ *is equal to zero*; for an even $\nu(= 2\mu$, say) however, we have

$$G_{2\mu-2\rho} = c^{\mu-\rho} \frac{(2\mu - 2\rho)!}{2^{\mu-\rho}(\mu - \rho)!}$$

and we obtain

(38) $$\begin{aligned} \lim_{n\to\infty} \frac{M_{2\mu}}{a^\mu} &= \sum_{\rho=0}^{\mu} \frac{(2\mu)!}{(2\mu - 2\rho)!2^\rho \rho!} \cdot \frac{(2\mu - 2\rho)!}{2^{\mu-\rho}(\mu - \rho)!} \\ &= \frac{(2\mu)!}{2^\mu \mu!} \sum_{\rho=0}^{\mu} \frac{\mu!}{\rho!(\mu - \rho)!} c^{\mu-\rho} = H_{2\mu}(1 + c)^\mu \end{aligned}$$

in accordance with (36). Lemma III is therefore proved.

Our principal theorem is now an obvious consequence of the well known theorem of the continuity of moments. By virtue of this theorem the convergence of $V_n(a_n + ys_n\sqrt{2})$ towards a normal distribution as given by (7) will indeed be assured if the moments of V_n converge towards the moments of the corresponding normal distribution; i.e. if (34) is true. Thus our principal theorem is completely demonstrated.

4. Some applications.

EXAMPLE 1. We shall consider the following play as a very simple application of our theorem: An urn contains $m = 2n$ counters bearing the numbers 1, 2, $\cdots$ m. We draw them all, one after the other, without returning the counters previously drawn. *We ask for the probability $P_{2n}(x)$ that an even counter will appear at a drawing of even number x times* ($0 \leqq x \leqq n$).
As can be easily found, we have

$$p_2 = p_4 = \cdots = p_{2n} = \tfrac{1}{2}$$

$$p_{2,4} = p_{2,6} = \cdots = p_{2n-2,2n} = \frac{1}{4}\frac{2n-2}{2n-1}$$

Consequently

$$(39)\quad \begin{aligned} &S_1 = \frac{n}{2}, \qquad S_2 = \binom{n}{2}\frac{1}{4}\frac{2n-2}{2n-1}, \qquad S_3 = \binom{n}{3}\frac{1}{8}\frac{(2n-2)(2n-4)}{(2n-1)(2n-2)}, \\ &\cdots\cdot \\ &S_z = \frac{1}{2^z}\binom{n}{z}\frac{(2n-2)(2n-4)\cdots(2n-2z+2)}{(2n-1)(2n-2)\cdots(2n-z+1)}. \end{aligned}$$

From (39) it follows that

$$(40)\qquad g_n(z) = \frac{n-z}{n}\frac{2n-2z}{2n-z}.$$

Setting $z/\frac{1}{2}n = u$, we get

$$(41)\qquad h_n(u) = \frac{(2-u)^2}{2\left(2-\dfrac{u}{2}\right)}.$$

The conditions (i), (ii), (iii) of our principal theorem are obviously satisfied if $\epsilon < 4$ and we have

$$h_n'(0) = -\tfrac{3}{4} = c$$

The probability defined above is thus seen to converge (according to (11)) *towards a normal distribution, having a mean equal to* $\dfrac{n}{2}$ *and a variance* $M_2 \sim \dfrac{n}{8}$

EXAMPLE 2. *Probability of an "occupation."* Let k stones be distributed by chance over n places. Then the probability that any stone will occupy a certain place will be equal to $1/n$. *We ask for the probability $P_n(x)$ that there shall be x places, every one of which is occupied by exactly m stones.*[4]

By certain simple considerations, well known in combinatory calculus, we obtain:

$$(42)\qquad a_n = n\frac{k!}{m!(k-m)!}\left(\frac{1}{n}\right)^m\left(1-\frac{1}{n}\right)^{k-m}$$

---

[4] The problem presents itself for instance if we ask for the probability that in a certain county there will be x villages, everyone of m inhabitants.

$$S_z = \frac{n!}{z!(n-z)!}\,\frac{k!}{(m!)^z(k-mz)!}\left(\frac{1}{n}\right)^{mz}\left(1-\frac{z}{n}\right)^{n-mz}. \tag{43}$$

Let $k/n = \alpha$. From (43) we deduce that

$$\begin{aligned} g_n(z) = \frac{n-z}{n}\left(\frac{1-\dfrac{z+1}{n}}{1-\dfrac{z}{n}}\right)^{n\alpha} \frac{1}{\left(1-\dfrac{1}{n}\right)^{n\alpha}}\,\frac{\left(1-\dfrac{z}{n}\right)^{mz}\left(1-\dfrac{1}{n}\right)^{m}}{\left(1-\dfrac{z+1}{n}\right)^{mz+m}} \\ \cdot\frac{\left(\alpha-\dfrac{zm}{n}\right)\left(\alpha-\dfrac{zm+1}{n}\right)\cdots\left(\alpha-\dfrac{zm+m-1}{n}\right)}{\alpha\left(\alpha-\dfrac{1}{n}\right)\cdots\left(\alpha-\dfrac{m-1}{n}\right)} \end{aligned} \tag{44}$$

Now, let n *and* k *tend simultaneously to* ∞, *in such a way that* $\alpha = \dfrac{k}{n}$ *remains bounded.* We get at first

$$\lim_{n\to\infty}\frac{a_n}{n} = \frac{\alpha^m}{m!}e^{-\alpha}. \tag{45}$$

As a_n is seen to be of the order of magnitude of n we introduce the new variables

$$\frac{z}{n} = v \quad\text{and}\quad v = u\,\frac{a_n}{n}.$$

We have then (writing h and $\bar{h}$ instead of h_n and $\bar{h}_n$):

$$g_n(z) = g_n(nv) = \bar{h}(v)$$

$$\bar{h}(v) = \bar{h}\left(u\,\frac{a_n}{n}\right) = h(u).$$

Therefore

$$\begin{aligned} \bar{h}(v) = (1-v)\left(1-\frac{1}{n(1-v)}\right)^{n\alpha}\frac{1}{\left(1-\dfrac{1}{n}\right)^{n\alpha}}\left(\frac{1-\dfrac{1}{n}}{1-v-\dfrac{1}{n}}\right)^{m} \\ \left(1-\frac{1}{n(1-v)}\right)^{-nmv}\cdot\frac{(\alpha-mv)\left(\alpha-\dfrac{1}{n}-mv\right)\cdots\left(\alpha-\dfrac{m-1}{n}-mv\right)}{\alpha\left(\alpha-\dfrac{1}{n}\right)\cdots\left(\alpha-\dfrac{m-1}{n}\right)}. \end{aligned} \tag{46}$$

These formulae show that the k-th derivative of $\bar{h}(v)$ with respect to v contains only rational expressions, [in the denominators of which there appear powers of $(1-v)$], and positive powers of $\log\left(1-\dfrac{1}{n(1-v)}\right)$. The conditions (i) and (iii) of our principal theorem are therefore satisfied if $\epsilon < 1$. Furthermore we have

$$\left(\frac{d\bar{h}}{dv}\right)_{v=0} = -1 - \frac{\alpha}{1 - \frac{1}{n}} - mn \log\left(1 - \frac{1}{n}\right) + \frac{m}{1 - \frac{1}{n}}$$

$$(47) \qquad -m \frac{\left(\alpha - \frac{1}{n}\right)\left(\alpha - \frac{2}{n}\right) \cdots \left(\alpha - \frac{m-1}{n}\right) + \alpha\left(\alpha - \frac{2}{n}\right) \cdots \left(\alpha - \frac{m-1}{n}\right) + \cdots \alpha\left(\alpha - \frac{1}{n}\right) \cdots \left(\alpha - \frac{m-2}{n}\right)}{\alpha\left(\alpha - \frac{1}{n}\right) \cdots \left(\alpha - \frac{m-1}{n}\right)}$$

and consequently

$$\lim_{n\to\infty} \left(\frac{dh}{du}\right)_{u=0} = \left[-1 - \frac{m^2}{\alpha} - \alpha + 2m\right] \lim_{n\to\infty} \frac{a_n}{n}$$
$$= -\left(1 + \frac{(m-\alpha)^2}{\alpha}\right) \frac{\alpha^m}{m!} e^{-\alpha} = c.$$

We have thus obtained the interesting result that,

The probability $V_n(x)$ that x places at most are occupied, each one by m stones, converges towards a normal distribution if k and n tend simultaneously to ∞ in such a way that $\lim_{n\to\infty} \frac{k}{n} = \alpha$ is bounded. We have then

$$(48) \qquad \lim_{n\to\infty} V_n(a_n + u\sqrt{2}\, s_n) = \phi(u)$$

with

$$(49) \qquad \lim_{n\to\infty} \frac{a_n}{n} = \frac{\alpha^m}{m!} e^{-\alpha}, \qquad \lim_{n\to\infty} \frac{s_n^2}{a_n} = 1 - \frac{\alpha^m e^{-\alpha}}{m!} \cdot \left[1 + \frac{(m-\alpha)^2}{\alpha}\right].$$

University of Istanbul,
Istanbul, Turkey.

REFERENCES

[1] H. Poincaré, *Calcul des Probabilités*, Paris, 1912
[2] W. Burnside, *Theory of Probability*, Cambridge, 1928
[3] G. U. Yule, *Introduction to the Theory of Statistics*, London, 1932
[4] C. Jordan, *Acta Litter, Scient.* (Szégéd), Vol. III (1927), p. 193
[5] C. Jordan, *Acta Litter. Scient.* (Szégéd), Vol. VII (1934), p. 103
[6] E. J. Gumbel, *Comptes Rendus*, Vol. 202, (1936), p. 1627
[7] E. J. Gumbel, *Giorn. Inst. Ital. Att.*, Vol. XVI (1938)
[8] H. Geiringer, *Comptes Rendus*, Vol. 204 (1937), p. 1856
[9] H. Geiringer, *Comptes Rendus*, Vol. 204 (1937), p. 1914
[10] H. Geiringer, *Revue Interbalconique*, (Athens), Vol. II, pp. 1-26
[11] R. von Mises, *Zs. Aug. Math. und Mech.*, Vol. I (1921)

FIDUCIAL DISTRIBUTIONS IN FIDUCIAL INFERENCE*

By S. S. Wilks

1. **Introduction.** The essential idea involved in the method of argument now known as fiducial argument, at least in a very special case, seems to have been introduced into statistical literature by E. B. Wilson [1] in connection with the problem of inferring, from an observed relative frequency in a large sample, the true proportion or probability p associated with a given attribute. Since 1930 the ideas and terminology surrounding the fiducial method have been developed by R. A. Fisher [2, 3], J. Neyman [4, 5] and others into a system for making inferences from a sample of observations about the values of parameters which characterize the distribution of the hypothetical population from which the sample is assumed to have been drawn. The functional form of the population distribution law is assumed to be known The parameters may be means, a difference between means, variances, ranges, regression coefficients, probabilities or any other descriptive indices or combinations of indices which may be considered important in specifying the distribution function of a population. In arguing fiducially about the value of a parameter, a procedure applicable to some of the simple cases begins by the calculation from the sample of an *estimate* of the parameter in question. The values of the estimate in repeated samples of the same size will theoretically cluster "near" the true value of the parameter according to a certain distribution law which can, in general, be deduced from the functional form of the population distribution law. If the distribution of the estimate involves only the one parameter, and if, as is frequently the case, one can find a function ψ of the estimate and the parameter which has a distribution not depending on the parameter, then one is able to set up, in a rather simple manner, *fiducial limits* or a *confidence interval* for the parameter corresponding to the observed value of the estimate. The limits will depend on the particular method of calculating the estimate, the value of the estimate in the sample, and on the degree of risk of being wrong which one is willing to take in stating that the limits will include between them the value of the parameter for the population under consideration. In general the smaller the degree of risk, the wider apart will be the limits. Thus for a given pair of limits there will be an associated degree of uncertainty that the true value of the parameter is actually included between those limits. This uncertainty can be expressed by a probability α calculated from the sampling distribution of the ψ function of the parameter and estimate. Under certain conditions, one can, by simply changing variables, obtain from the ψ

* An expository paper presented to the American Statistical Association on December 28, 1937, at the invitation of the Program Committee.

distribution what has been termed by Fisher a *fiducial distribution* function of the parameter. From the fiducial distribution and for a given value of the estimate one can actually determine fiducial limits of the parameter corresponding to a given risk α. It will be seen as we proceed that the fiducial distribution plays no indispensable part in fiducial inference; the ψ function and its distribution from which the fiducial distribution is derivable, are sufficient for the fiducial argument in many cases that commonly arise in statistics. We shall discuss fiducial argument and fiducial distributions from the point of view of ψ functions.

2. **Example**. To illustrate these points let us consider an example, namely, the problem of determining fiducial limits and the fiducial distribution of the range of a rectangular distribution for a given value of the range in a sample "randomly drawn" from it.

If a sample of n individuals is drawn from a population whose distribution law is $f(x, \theta) = 1/\theta$, where only values of x between 0 and θ are considered, (that is, a rectangular distribution having range θ) the probability that the range r of the sample lies between r and $r + dr$ is $\varphi(r, \theta)\, dr$, where

$$\varphi(r, \theta) = \frac{n(n-1)}{\theta^n} (\theta - r)r^{n-2}. \tag{1}$$

Here θ is the parameter under question, and r is the estimate; r is the difference between the largest and smallest variate in the sample. Thus, for a given value of θ, say θ_0, $\varphi(r, \theta_0)$ is a sampling distribution law defined for given values of r on the range $r = 0$, to $r = \theta_0$. If we let $r/\theta = \psi$, then

$$\varphi(r, \theta)\, dr = n(n-1)(1-\psi)\psi^{n-2}\, d\psi = G(\psi)\, d\psi, \tag{2}$$

which, from a statistical point of view, shows that if we should take an aggregate of randomly drawn samples (of n items each) from rectangular populations and calculate ψ for each *sample-population combination*, then the distribution of ψ will be that given in (2). By a *sample-population combination* in this example we mean any rectangular population that may arise and a "randomly drawn" sample from it. The possible values of ψ range from 0 to 1. Thus if ψ_α is such that

$$n(n-1)\int_0^{\psi_\alpha} (1-\psi)\psi^{n-2}\, d\psi = \alpha, \quad \text{i.e.} \quad \psi_\alpha^{n-1}[n - (n-1)\psi_\alpha] = \alpha, \tag{3}$$

and if we draw a sample of n from a rectangular population, we can claim that the probability is $1 - \alpha$ that the ψ produced by this sample-population combination will satisfy the inequality

$$\psi_\alpha < \psi < 1. \tag{4}$$

It should be observed that there are many pairs of numbers, say ψ'_α and ψ''_α such that we can claim that $\psi'_\alpha < \psi < \psi''_\alpha$, with probability $1 - \alpha$ of being

correct in making the claim. ψ'_α and ψ''_α are ordinarily chosen so that the interval formed by them is as short as possible (or approximately so) in some sense. Inequality (4) is equivalent to each of the following inequalities

$$(5) \qquad \psi_\alpha < \frac{r}{\theta} < 1, \qquad \frac{r}{\psi_\alpha} > \theta > r.$$

Now ψ_α can be determined from (3) when n and α are given. For example, if $\alpha = .01$ and $n = 10$, we find from (3) that $\psi_\alpha = .495$. For a given sample, the fiducial limits r/ψ_α and r can be calculated from ψ_α and the sample. It will be noticed that fiducial limits are nothing more nor less than random variables that fluctuate from sample to sample. The interval between r and r/ψ_α is called a *confidence interval* or *fiducial interval*; $1 - \alpha$ is known as the *confidence coefficient* [4] associated with the limits. Hence, in repeated samples of n from a rectangular population with range θ_0, $100(1 - \alpha)$ percent of the samples will produce fiducial limits r/ψ_α and r which include the fixed value θ_0 between them. This statement holds regardless of the value of θ_0. Hence in an aggregate of sample-population combinations, the aggregate of pairs of fiducial limits r/ψ_α and r will, in $100(1 - \alpha)$ percent of the combinations, include between them the true value of the range of the population. Furthermore, whether there is only one rectangular population for all sample-population combinations or many different rectangular populations, this statement remains true, thus showing that the method of fiducial limits for inferring the value of the parameter is independent of any *a priori* distribution of rectangular populations in an aggregate of sample-population combinations—the distribution being with respect to values of θ.

Let us look at the matter geometrically. Suppose we are drawing samples from a rectangular population with $\theta = \theta_0$. The r for each sample is represented by a dot along Or in Figure 1; corresponding to each dot there is confidence interval cutting across the V-shaped region MOR. The probability is $1 - \alpha$ that a confidence interval computed from a sample from the population having range θ_0 will cut the line $\theta_0 K$. The cutting of $\theta_0 K$ by a confidence interval is equivalent to the statement that θ_0 is included between the corresponding fiducial limits.

From a practical statistical point of view what we have said has the following meaning: If on each occasion in which a randomly drawn sample of n from some rectangular population is considered, one (i) calculates the numbers r/ψ_α and r, and (ii) asserts that the range in the population producing the sample lies between these two computed limits, then in about $100(1 - \alpha)$ percent of the cases assertion (ii) will be correct (theoretically). Thus, in dealing with samples of 10 individuals from rectangular populations, one would be correct (theoretically) in about 99 percent of the cases by asserting that the population range will lie between the sample range and $2.020 \left(= \frac{1}{.495}\right)$ times the sample range. More generally, one need not use the same value of n all the way

through, provided that for the given α one evaluates ψ_α according to (3), for each n that arises. It will be seen from (3) that as n increases, the value of ψ_α tends to 1 and hence the fiducial limits r/ψ_α and r for any given sample tend to the same value, namely the sample range, thus showing that fiducial inferences about θ can be made arbitrarily certain by taking sufficiently large samples.

It is evident that the method of fiducial limits furnishes a satisfactory procedure for inferring the value of the population range θ from samples drawn from rectangular populations. Let us now go a step further and consider the fiducial distribution of θ and how it fits into the scene. The cumulative distribution of ψ is

$$\psi^{n-1}\,[n - (n-1)\psi] \tag{6}$$

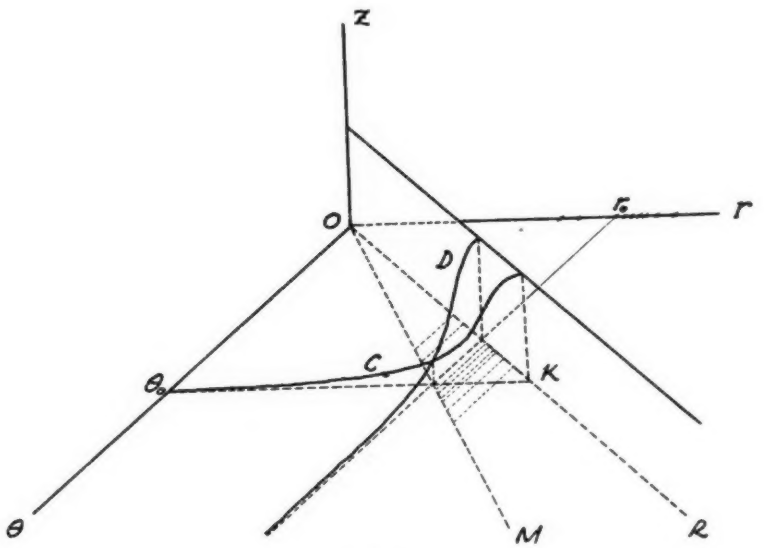

FIG. 1

and hence the cumulative distribution of r for a fixed θ, say θ_0, is

$$F(r, \theta_0) = \left(\frac{r}{\theta_0}\right)^{n-1}\left[n - (n-1)\left(\frac{r}{\theta_0}\right)\right] \tag{7}$$

which increases from 0 to 1 as r increases from 0 to θ_0. Geometrically, $z = F(r, \theta)$ can be represented as a surface defined over the region bounded by lines $O\theta$ and OR in Figure 1, such that z is zero along O and is unity along the line OR $(r = \theta)$. $F(r, \theta)$ is continuous inside the region θOR, and for any given value $r_0 \neq 0$ of r, $F(r, \theta)$ decreases from 1 to 0 as θ increases from r_0 to ∞. The curves having the equations

$$\begin{cases} z = F(r, \theta) \\ \theta = \theta_0 \end{cases} \quad \text{and} \quad \begin{cases} z = F(r, \theta) \\ r = r_0 \end{cases}$$

(where θ_0, r_0, and α are such that $r_0/\theta_0 = \psi_\alpha$ and $F(r_0, \theta_0) = \alpha$) are the curves C and D respectively. C is the cumulative distribution of ranges of samples of n from a rectangular population with range θ_0. The curve D has the mathematical characteristics of a cumulative distribution function cumulated in the negative direction with respect to θ: its ordinates increase from 0 to 1 as θ decreases from ∞ to θ_0. Thus, if we take $-\frac{\partial}{\partial\theta} F(r_0, \theta)$ we get a function $g(\theta, r_0)$ which has the essential mathematical characteristics of a distribution function: it is non-negative, can be integrated over any interval of θ, and has total area under it equal to unity. We have

$$(8) \qquad g(\theta, r_0) = n(n-1)\frac{r_0^{n-1}}{\theta^n}\left(1 - \frac{r_0}{\theta}\right)$$

and it is called the *fiducial distribution* of θ for $r = r_0$. It must be firmly pointed out that θ is not a random variable and hence $g(\theta, r_0)$ is *not* a distribution function of a random variable, although it has the mathematical properties of such a distribution. Objections have been raised to the use of the term fiducial distribution on the grounds that the thing to which it applies is not a distribution at all. However, as long as the term is carefully defined there should be no ambiguity in using it. From an analytical point of view, the problem of obtaining the fiducial distribution of θ is only a matter of changing variables for since

$$(9) \qquad \varphi(r, \theta)\, dr = g(\theta, r)\, d\theta = n(n-1)(1-\psi)\psi^{n-2}\, d\psi$$

and $\psi_\alpha = r_0/\psi_0$, we have

$$(10) \quad \int_{\theta_0\psi_\alpha}^{\theta_0} \varphi(r, \theta_0)\, dr = \int_{r_0}^{r_0/\psi_0} g(\theta, r_0)\, d\theta = \int_{\psi_\alpha}^{1} n(n-1)(1-\psi)\psi^{n-2} d\psi = 1 - \alpha.$$

We remark again that

$$(11) \qquad \int_{r_0}^{\theta_1} g(\theta, r_0)\, d\theta$$

is *not* to be interpreted as probability as though θ were a random variable. Instead, the meaning is as follows: Let r_0 be the range in a sample known to be from *some* rectangular population, and let the value of r_0 be inserted in (11), and let θ_1 be determined so that the value of the integral is $1 - \alpha$. The two limits for the integral are fiducial limits associated with the sample for the confidence coefficient $1 - \alpha$, which were discussed earlier. Thus, for each sample, we can compute fiducial limits using the fiducial distribution. These limits, as we have seen by considering the ψ function, fluctuate from sample to sample in such a way that the probability is $1 - \alpha$ that they will include between them the true value of the range of the population under consideration.

3. **Summary of Principles**. From the point of view we have taken the essential notions involved in the method of fiducial argument and fiducial

distributions for the case of a continuous variate and one parameter can be readily abstracted from the example just discussed. In general, we have the following steps:

(a) A sample is assumed to be *randomly drawn* from a population with a distribution of *known* functional form $f(x, \theta)$, θ being a parameter. Let $x_1, x_2, \cdots, x_n$ be the values of x in the sample.

(b) A function, say $\psi(x_1, \cdots, x_n, \theta)$ of the sample x's and θ is devised so that its sampling distribution $G(\psi)$ involves θ and the x's only as they enter into ψ. The value of θ in ψ is that for the population from which the sample is actually drawn.

(c) Two numerical values of ψ, say ψ'_α and ψ''_α are chosen (ordinarily as close together as possible) so that the probability computed from $G(\psi)$ is $1 - \alpha$ (e.g. 0.95) that ψ will lie between ψ'_α and ψ''_α—more briefly $P(\psi'_\alpha < \psi < \psi''_\alpha) = 1 - \alpha$.

(d) The inequality $\psi'_\alpha < \psi < \psi''_\alpha$ which contains only one unknown, namely θ, is solved for θ giving the equivalent inequality $\underline{\theta} < \theta < \bar{\theta}$ where $\underline{\theta}$ and $\bar{\theta}$ are *fiducial limits* and are subject to sampling fluctuations.

(e) The expression $P(\psi'_\alpha < \psi < \psi''_\alpha) = 1 - \alpha$ is replaced by the equivalent expression $P(\underline{\theta} < \theta < \bar{\theta}) = 1 - \alpha$ which states that the probability is $1 - \alpha$ that a sample will yield values $\underline{\theta}$ and $\bar{\theta}$ which will include the true value of θ between them.

(f) The differential element for the fiducial distribution of θ is $G(\psi) \left| \dfrac{\partial \psi}{\partial \theta} \right| d\theta$ (provided $\partial\psi/\partial\theta$ is a function of θ which does not change sign for a given sample of x's) and is obtained by letting θ be the variable in $G(\psi)\, d\psi$, keeping the x's fixed.

To give precisely the conditions under which all of these steps can be performed is a technical matter which will not be considered here. It is sufficient to remark that they can be performed in many cases of practical interest. Fiducial argument can be carried on using only the first five steps without introducing the notion of a fiducial distribution. In connection with step (a) it should be particularly noticed that the functional form $f(x, \theta)$ of the population under question is assumed to be known and that the sample under consideration is "randomly drawn" from the population. Thus, in applying the theory to practical problems it is a matter of fundamental importance that these two assumptions be valid. In cases where a sufficient amount of data exists, it can usually be satisfactorily tested by using the χ^2 test and other devices, whether or not a given functional form for $f(x, \theta)$ is a valid assumption. In cases where sufficient data do not exist for actually making such a test justification for assuming a given function form usually has to be made on the basis of theoretical considerations. From a practical point of view the notion of randomness is characterized by methods of drawing samples rather than *a posteriori* mathematical considerations of the sample after it has been drawn, and thus the question of randomly drawing samples depends largely upon the

experience and sound judgment of the experimenter. However, after one or more samples have been drawn "at random," the problem of arguing from them about the populations from which they were drawn is largely mathematical.

4. **Case of large samples**. For a population with a distribution of known functional form, a fiducial distribution of the parameter clearly depends on the size of the sample and the particular estimate used. For example, in large samples, we would get a fiducial distribution of the mean of a normal population of known variance by using the sample mean which would be different from the one obtained using the median of the sample. In order to be able to make the inferences about θ as accurate as possible, a ψ function should theoretically be used which will produce fiducial limits which are closest together, on the average, or perhaps "best" in some other sense, for a given α. The fiducial distribution obtainable from such a ψ could then be referred to as the "best" fiducial distribution, and theoretically it should be used in preference to other possible fiducial distributions if fiducial distributions are to be used at all to set fiducial limits. In large samples from a population with a distribution function $f(x, \theta)$, it is known [6] that, under rather general conditions, fiducial limits which are closest together on the average can be obtained by letting

$$\psi = \frac{1}{\sqrt{n}}\left(\frac{\partial L}{\partial \theta}\right)\left[E\left\{\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right\}\right]^{-\frac{1}{2}} \tag{12}$$

and treating ψ as a normally distributed variate with zero mean and unit variance, where $L = \sum_{i=1}^{n} \log f(x_i, \theta)$, the logarithm of the likelihood of θ for the given sample, $x_1, x_2, \cdots, x_n$ are values of x in the sample, and E denotes mathematical expectation. For example, in the case of a binomial population where each individual belongs either to class A or class B, we have $f(x, \theta) = \theta^x(1 - \theta)^{1-x}$ where θ is the probability associated with class A, x will be 0 or 1 according to whether an individual belongs to B or A. In a sample of n individuals, $L = m \log \theta + (n - m) \log (1 - \theta)$, where m is the number of individuals in class A. $E\left\{\left(\frac{\partial \log f(x, \theta)}{\partial \theta}\right)^2\right\} = \frac{1}{\theta(1 - \theta)}$, and we get $\psi = \frac{m - n\theta}{\sqrt{n\theta(1 - \theta)}}$. If we should want to find fiducial limits of θ for a confidence coefficient of .95, we would solve (1) the equations $\frac{m - n\theta}{\sqrt{n\theta(1 - \theta)}} = \pm 1.96$ for θ, thus getting two values of θ, say $\underline{\theta}$ and $\bar{\theta}$. We can then say that $\underline{\theta}$ and $\bar{\theta}$ will include the true value of θ between them with a probability of .95 of being correct, in the sense that if we applied this rule consistently to samples from binomial populations, we would have a procedure that would lead to a correct statement in about 95 percent of the cases (theoretically).

To illustrate the difference between the fiducial method and the commonly

used method of placing limits on θ for $P = .95$, consider an example in which $m = 150$, $n = 400$. The usual procedure is to replace θ by m/n in $\theta \pm 1.96 \sqrt{\frac{\theta(1-\theta)}{n}}$, which yields .311 and .431. The fiducial procedure is to solve the equation $\frac{m - n\theta}{\sqrt{n\theta(1-\theta)}} = \pm 1.96$, for θ, thus obtaining .312 and .455. For the case of small samples, the problem of getting "best" fiducial limits becomes more complicated [5].

5. **Extensions of Fiducial Argument.** It will be observed that it is not necessary for ψ to be a function of only one statistic and θ in order to be able to argue fiducially about θ. For example, if a sample of n is drawn from a normal population with mean θ, it is well known that if $\bar{x}$ is the sample mean then

$$\psi = \frac{(\bar{x} - \theta)\sqrt{n(n-1)}}{\left[\sum_1^n (x_i - \bar{x})^2\right]^{\frac{1}{2}}} \tag{13}$$

(which is Fisher's t function), has the distribution

$$\frac{\Gamma(\frac{1}{2}n)}{\Gamma(\frac{1}{2}(n-1))\sqrt{\pi(n-1)}} \frac{d\psi}{[1 + \psi^2/(n-1)]^{\frac{1}{2}n}}. \tag{14}$$

Here ψ is a function of two statistics, namely $\bar{x}$ and $\sum_{i=1}^{n} (x_i - \bar{x})^2$, and the fiducial distribution of θ for this ψ function is obtained at once by applying rule (*f*).

The ideas of fiducial argument may be extended in other directions, but these cannot be considered in any detail here. For example ψ may be a function of $x_1, \cdots, x_n$ and two or more population parameters, in which case one could set up fiducial *regions* for the several parameters. From a practical point of view, the fiducial argument for two or more parameters simultaneously, had hardly been touched. Again ψ may be a function of statistics from two samples, one observed and the other not yet observed, and not involving population parameters, at all, in which case one can argue fiducially about the statistic in question for the unobserved sample [3]. The notion of a fiducial distribution has been extended to several parameters taken simultaneously [3, 7], but the problem of working out relations between fiducial distributions of several parameters and fiducial regions is yet to be investigated. The principles may be readily applied in situations in which the x's involved in ψ take on discrete values. In this case the equality signs in the probability expressions in steps (c) and (d) would be replaced by greater than or equal signs ($\geq$). Two excellent examples of the application of principles of fiducial argument to the discrete case are furnished: (i) by a paper by Pearson and Clopper [8] on fiducial limits of the probability P from samples from a binomial population, and (ii) by a paper by Ricker [9] on fiducial limits of m in the Poisson distribution $f(x, m) = m^x e^{-m}/x!$.

REFERENCES

[1] E. B. Wilson, "Probable Inference, the Law of Succession, and Statistical Inference," *Jour. American Statistical Association*, Vol. 22 (1927), pp. 209-212.

[2] R. A. Fisher, "The Concepts of Inverse Probability and Fiducial Probability Referring to Unknown Parameters," *Proc. Royal Society of London*, Series A, Vol. 139 (1933), pp. 343-348.

[3] R. A. Fisher, "The Fiducial Argument in Statistical Inferences," *Annals of Eugenics*, vol. 6 (1935), pp. 391-398.

[4] J. Neyman, "On the Two Different Aspects of the Representative Method: the Method of Stratified Sampling and the Method of Purposive Selection," *Royal Statistical Society*, vol. 97, 1934, pp. 558-625.

[5] J. Neyman, "Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability," *Phil. Trans. Roy. Soc. London*, Series A, Vol. 236 (1937), pp. 333-380.

[6] S. S. Wilks, "Shortest Average Confidence Intervals from Large Samples," *Annals of Mathematical Statistics*, Vol. 9 (1938), pp. 166-175.

[7] I. E. Segal, "Fiducial Distribution of Several Parameters with Application to a Normal System," *Proc. Cambridge Phil. Soc.*, vol. 34 (1938), pp. 41-47.

[8] C. J. Clopper and E. S. Pearson, "The Use of Confidence or Fiducial Limits in the Case of the Binomial," *Biometrika*, vol. 24, 1934, pp. 404-413.

[9] William E. Ricker, "Fiducial Limits of the Poisson Frequency Distribution," *Jour. American Statistical Association*, vol. 32, 1937, pp. 349-356.

Princeton University.

BIOLOGICAL APPLICATIONS OF NORMAL RANGE AND ASSOCIATED SIGNIFICANCE TESTS IN IGNORANCE OF ORIGINAL DISTRIBUTION FORMS*

By William R. Thompson

The word *normal* has been used in many senses—commonly by statisticians to designate a well-known distribution function. Another use familiar to biologists, particularly in experimental work and medicine, is to denote an untreated or control part of a universe, or a part whose members are free from specified characteristics such as evidence of past or present disease or malformation. Closely related to this last usage are attempts to delimit so-called normal ranges of variation for a quantitative attribute of the members of part or all of a universe in question. Interpretations are often vague, as when the interval between the least and greatest values observed in either a large or a small number of instances is taken to estimate a normal range. We shall consider the problem of using ranked data for estimating normal ranges as defined in the next paragraph.

If the instances have been drawn at random from a universe (U) of all possible observations obtainable in a prescribed manner, and are enumerated in ascending order of magnitude, $\{x_i\}$ for $i = 1, \cdots, n$; then it is proposed to show in the present communication how ranges of the type (x_k, x_{n+1-k}) may be used to estimate *normal ranges*, R_f, where the subscript f is the theoretical probability that a random value, x, drawn from U will lie within the range R_f, g that it will lie above, and g that it will lie below (where $2g = 1 - f$). Furthermore, it is proposed to show how these ranges may be used as the basis of significance tests where altered conditions appear to lead to abnormal biological variation. The form of frequency-distribution of U is supposed unknown, and is without effect upon the analysis. *Section* 1 is a development of the theory of range estimation, treated briefly in a previous paper [1] together with illustrations of its application. *Section* 2 deals with significance tests.

1. **The Method of Range Estimation**. Let x be a real variate, a random value drawn from an infinite universe or population U. Let $f(x)$ be the frequency function of x in U, supposed unknown; and $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Then for any given α and β, where $\alpha < \beta$, and

$$P(\alpha < x < \beta) \equiv \int_{\alpha}^{\beta} f(x)\,dx. \tag{1}$$

* Presented at a meeting of the American Statistical Association, December 28, 1937, Atlantic City, N. J.

To facilitate development, suppose that in any finite sampling under consideration no two values of x may be exactly the same. Let $S = \{x_k\}$, $k = 1, \cdots, n$, denote a random sample from U, where the order of enumeration is arbitrary, but temporarily taken as a random order (to fix the ideas, consider this the order obtained in drawing). Let p_k be defined by

$$(2) \qquad p_k = P(x < x_k) \equiv \int_{-\infty}^{x_k} f(x)\,dx \quad \text{from which} \quad dp_k = f(x_k)\,dx_k .$$

Then p_k is the probability that a random x from U shall be less than any number x_k. Then obviously if x_k is drawn at random from U, p_k is a random variable whose distribution is the unit rectangle; i.e., $P(p' < p_k < p'') = p'' - p'$. Furthermore, the joint probability that x_k will lie in the interval x_k, $x_k + dx_k$ and that exactly r values in the sample S will be less than x_k is, to within terms of order dp_k, $\binom{n-1}{r} \cdot p_k^r \cdot (1 - p_k)^{n-1-r}\,dp_k$.

Then, in repeated sampling as above, for the case where just r of the n random values $\{x_i\}$ are less than the k-th drawn, let $P_{n,r}(p' < p_k < p'')$ denote the probability that p_k lies in the interval (p', p''). Then

$$(3) \qquad P_{n,r}(p' < p_k < p'') = \frac{(r + s + 1)!}{r! \cdot s!} \cdot \int_{p'}^{p''} p^r \cdot q^s \cdot dp,$$

where $s = n - 1 - r$, and $q = 1 - p$. Obviously, the expression on the right of (3) does not depend on k if this index is the order of draft or a random index, but only upon the condition that exactly r of the n random values from U be less than a value x_k drawn at random from the sample of n values. Accordingly, we obtain the same result if we enumerate the n values $\{x_i\}$ in ascending order of magnitude ($x_i < x_j$, if $i < j$). Then $k = r + 1$, in the cases considered, and (3) may be written,

$$(4) \qquad P_n(p' < p_k < p'') = \frac{n!}{(k-1)!(n-k)!} \cdot \int_{p'}^{p''} p^{k-1} \cdot q^{n-k} \cdot dp,$$

for $0 \leqq p' \leqq p'' \leqq 1$. Obviously, the result is the same if we deal instead with the k-th value (x_k) of every random sample S drawn. In passing it may be noted that for $p' = 0$ and $p'' = p$ in (4) we have

$$(5) \qquad P_n(p_k < p) = I_p(k, n - k + 1),$$

which may be evaluated for k, $n - k + 1 \leqq 50$ by means of the *Tables of the Incomplete Beta-Function* [2].

Of course, $P_n(0 < p_k < 1) = 1$, and (4) gives $\bar{p}_k$, the mean value of p_k in repeated random sampling of n values from U, as

$$(6) \qquad \bar{p}_k = \frac{n!}{(k-1)!(n-k)!} \cdot \int_0^1 p^k \cdot q^{n-k} \cdot dp = \frac{k}{n+1} .$$

Similarly, the variance, $\sigma_{p_k}^2$, of p_k is given by

$$\sigma_{p_k}^2 = E[(p_k - \bar{p}_k)^2] = \frac{k(n-k+1)}{(n+1)^2\cdot(n+2)}. \tag{7}$$

Now suppose that we want to find a range (α, β) such that, in random drafts from U, the theoretical relative frequency of drawing x less than α is g, and the same as that of drawing x greater than β. (α, β) may be called a *central confidence range* with a *confidence* $f = 1 - 2g$ that x drawn at random from U will lie within the range. For $g = k/(n+1)$ we may take the range $R_f = (x_k, x_{n-k+1})$; and likewise with $g = 5\%$ we may estimate, or approximate by interpolation where $20k > n + 1 > 20(k-1)$, a range R_f for normal biological variation of a specified character, and this may be called briefly the estimated 90% *central normal range*.

Of course the probability of drawing $x < \alpha$ is $\int_{-\infty}^{\alpha} f(x)\,dx$, and that of drawing $x > \beta$ is $\int_{\beta}^{\infty} f(x)\,dx$; and these probabilities are unknown, as the frequency function $f(x)$ is unknown; but with $\alpha = x_k$ and $\beta = x_{n-k+1}$ the theoretical relative frequency in each case is $k/(n+1)$ regardless of the universe.

It has been shown [1] also that if the sample S were drawn at random from a finite ordered population of aggregate number N, denoted by U_N, and Np_k is the number of values in U_N that are less than the k-th member of the given random sample in ascending order of magnitude; then, for S a sample of n values as before, the mean value of p_k in repeated sampling is

$$\bar{p}_k = \frac{k}{n+1}\left(1 + \frac{1}{N}\right) - \frac{1}{N}, \quad \text{and}$$

$$\sigma_{p_k}^2 = \frac{k(n-k+1)}{(n+1)^2\cdot(n+2)}\cdot\left(1 + \frac{1}{N}\right)\left(1 - \frac{n}{N}\right).$$

An example is furnished by an analysis of data reported by Wadsworth and Hyman [3] in a study of influences of antigenic treatment of horses upon their plasma concentration of esterified cholesterol, free cholesterol, and phospholipids. As in chart 1 for normal horses, a graph has been constructed for each horse studied, using time as abscissa and a logarithmic ordinate scale for observed values of plasma concentration of the constituents:

1. Esterified Cholesterol,
2. Free Cholesterol, and
3. Phospholipids *times one-tenth*,

the respective successive points for each being joined to form three polygon curves. As these are in all cases discrete and lie in the order of enumeration from top to bottom of the graph, no special label seemed needed; but estimated normal ranges for the central 90% of variation have been indicated in each

case by two horizontal lines between brackets at the right, numbered to correspond with the enumeration above. The ranges are based on observations on 62 plasma samples, each from a different presumably normal horse. The normal horses in the chart show about the same individual variations; but, of course, the ranges are not to be interpreted to indicate normal variation for an individual animal.

Chart 2 presents in like manner the data obtained for horses under immunization against tetanus and the streptococcus. The tetanus immunization treat-

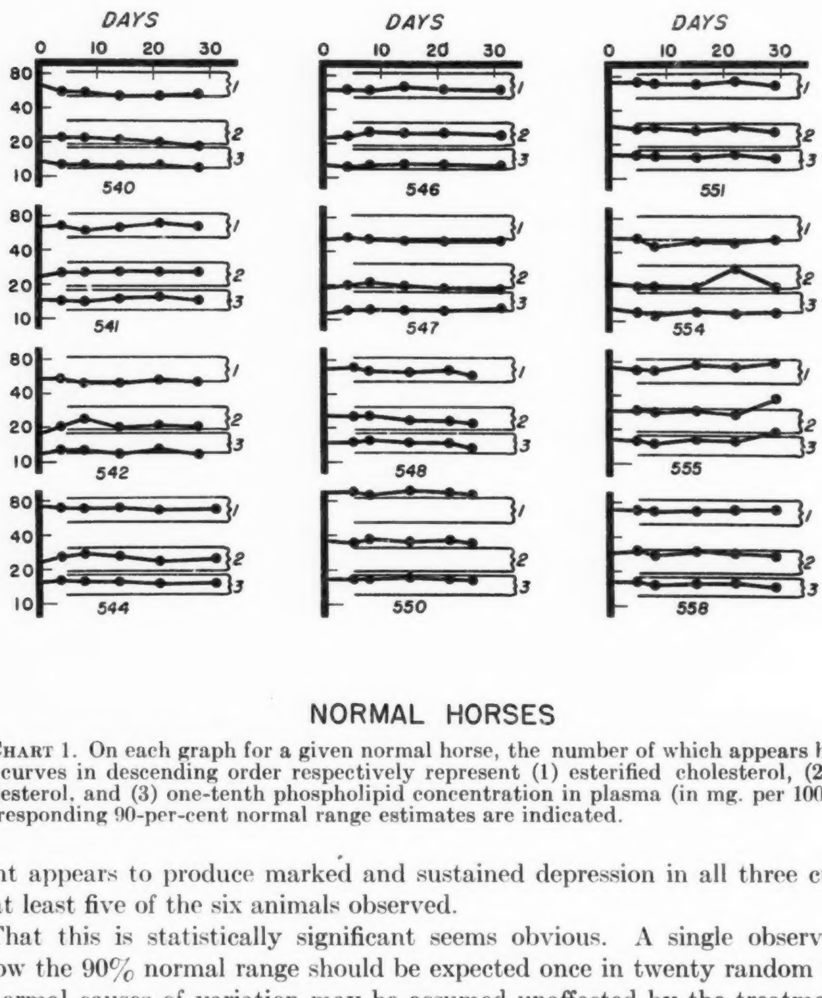

CHART 1. On each graph for a given normal horse, the number of which appears below, the curves in descending order respectively represent (1) esterified cholesterol, (2) free cholesterol, and (3) one-tenth phospholipid concentration in plasma (in mg. per 100 cc.). Corresponding 90-per-cent normal range estimates are indicated.

ment appears to produce marked and sustained depression in all three curves of at least five of the six animals observed.

That this is statistically significant seems obvious. A single observation below the 90% normal range should be expected once in twenty random trials if normal causes of variation may be assumed unaffected by the treatment in question. The expectation of obtaining 5 or more such values in six independent trials is obviously much less, and may be accurately estimated by means of relations developed in the following section.

2. **Significance Tests**. Now consider as in *section* 1 another sample S' of n' values; $\{x'_{k'}\}$, $k' = 1, \cdots, n'$ (where $x'_i < x'_j$ if $i < j$), drawn at random from an infinite universe U' as was S from U; but where U' and U are not necessarily

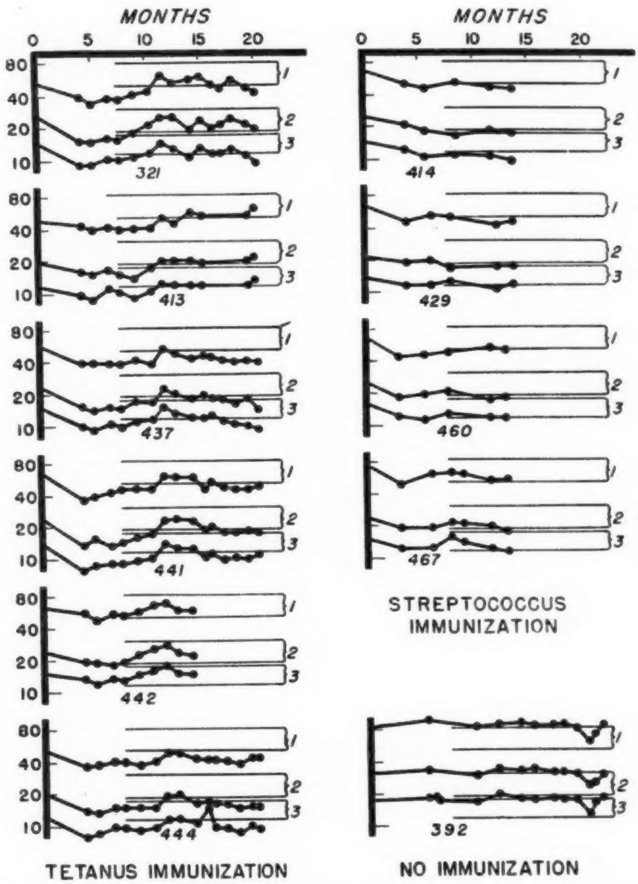

CHART 2. On each graph for horses receiving the indicated antigenic treatment and one untreated horse, the curves in descending order respectively represent (1) esterified cholesterol, (2) free cholesterol, and (3) one-tenth phospholipid concentration in plasma (in mg. per 100 cc.). Corresponding 90-per-cent normal range estimates are indicated.

the same universe. In like manner it may be shown that, if x' is drawn at random from U' and $p'_{k'}$ denotes $P(x' < x'_{k'})$, then

$$(8) \qquad P_{n'}(\phi' < p'_{k'} < \phi'') = \frac{(v + w + 1)!}{v!w!} \cdot \int_{\phi'}^{\phi''} p^v \cdot q^w \cdot dp$$

where $q = 1 - p$, $v = k' - 1$, $w = n' - k'$, and $0 \leqq \phi' \leqq \phi'' \leqq 1$.

The probabilities in (4) and (8) are independent, obviously, whether U' is the same as U or not. Accordingly, these relations make possible an evaluation

of $P(p_k < p'_{k'})$ under the circumstances where repeated sampling is applied to both the case of S and to that of S'. With this understanding, then

$$(9) \qquad P(p_k < p'_{k'}) = \frac{(r+s+1)!(v+w+1)!}{r!\cdot s!\cdot v!\cdot w!}\cdot \int_0^1 p_0^r\cdot q_0^s\cdot dp_0\cdot \int_{p_0}^1 p^v\cdot q^w\cdot dp,$$

where, as before, $r = k - 1$, $s = n - k$, $v = k' - 1$, $w = n' - k'$, $q \equiv 1 - p$, and $q_0 \equiv 1 - p_0$.

In a previous paper [4] a Ψ-function was defined as

$$(10) \qquad \Psi(r, s, r', s') \equiv \frac{\sum_{\alpha=0}^{r'} \binom{r + r' - \alpha}{r}\binom{s + s' + 1 + \alpha}{s}}{\binom{r + s + r' + s' + 2}{r + s + 1}}$$

for any four rational integers $r, s, r', s' \geqq 0$; and it was shown in detail that the right member of (9) is equal to $\Psi(r, s, v, w)$; whence we may write

$$(11) \qquad P(p_k < p'_{k'}) = \Psi(k - 1, n - k, k' - 1, n' - k').$$

Obviously, if U and U' are the same universe, then $p_k < p'_{k'}$ if and only if $x_k < x'_{k'}$, and then we have

$$(12) \qquad P(x_k < x'_{k'}) = \Psi(k - 1, n - k, k' - 1, n' - k')$$

in repeated random sampling applied to both sample types, S and S', respectively of n and of n' observations. In the paper just mentioned, and in another [5] the Ψ-function was further developed by extension of definition to include $\Psi(r, s, -1, s') \equiv 0$, and it was shown that

$$(13) \qquad \Psi(r, s, r', s') \equiv \Psi(r, r', s, s') \equiv \Psi(s', r', s, r) \equiv 1 - \Psi(s, r, s', r').$$

Further demonstrations [5] included the relation,

$$(14) \qquad \Psi(r, s, r', s') \equiv \frac{\sum_{\alpha=0}^{\alpha \leqq s, r'} \binom{r + r' + 1}{r + 1 + \alpha}\binom{s + s' + 1}{s - \alpha}}{\binom{r + s + r' + s' + 2}{r + s + 1}},$$

which offers another form for calculation. The identities in (13) are particularly useful to facilitate calculation where one of the four arguments is small. A system for forming a table has also been developed [4, 5] in an economical form, but tabulation has been given only for the arguments not exceeding 5.

Now, in applying a test based on relation (12) or on that for the complementary probability, $P(x'_{k'} < x_k)$ which obviously, by (13), equals $\Psi(n - k, k - 1, n' - k', k' - 1)$, we may wish to exclude from the *normal* set of observations those values obtained from animals later given the treatment in question in the statistical significance test. The purpose would be to avoid violation of the condition of independent sampling required. In the case of the tetanus antigen treatment, we have an experience wherein 5 or more of 6 horses treated yield

values for a given plasma constituent less than the third in ascending order of magnitude (namely x_3) in our independent set of *normal* values. Here $n' = 6$, and $n = 62 - 6 = 56$. In accordance with the hypothesis that the treatment in question does not affect normal causes of variation in the plasma constituent under investigation we have $P(x'_{k'} < x_3)$ is $\Psi(53, 2, 6 - k', k' - 1)$. This is approximately $1.891(10)^{-5}$ for $k' = 5$, and $4.555(10)^{-7}$ for $k' = 6$. Obviously, a rule for establishing the value of k to be used in such tests should be fixed in advance without prejudice, as in the present case where we have taken $k \geqq g(n + 1) > k - 1$ for $g = 5\%$.

In the case of streptococcus immunization treatment, the corresponding test would have $n = 58$, $n' = 4$, $k = 3$, and $k' = 4, 3$, or 2; which would yield approximately $2.689(10)^{-5}$, $1.031(10)^{-3}$, or $1.817(10)^{-2}$, respectively for $P(x'_{k'} < x_3)$. Thus it appears that where such values are found (intuitively it would appear a fortiori if we compare instead with x_3 of the entire normal set of 62 values), their low magnitude appears to discredit the hypothesis that such discrepancies are ascribable to mere chance normal variation in the quantitative attribute investigated.

The tests proposed are free from any assumption concerning the form of the original distribution $f(x)$. The illustrative material is only a part of that presented with similar statistical treatment in the paper of Wadsworth and Hyman [3], which makes it apparent that the tests suggested here may be useful and powerful in analysis of biological and other experimental data. From a similar point of view, Hotelling and Pabst [6] developed tests of bi-variate correlation, and Milton Friedman has elaborated a multi-variate rank analysis [7], the tests being likewise free from any assumption about the form of the original distributions. In a previous paper [1] confidence ranges for the median are based similarly, employing relation (5) for the special case $p = \frac{1}{2}$.

Division of Laboratories and Research
New York State Department of Health
Albany, N. Y.

REFERENCES

[1] W. R. Thompson, *Annals of Mathematical Statistics*, Vol. 7 (1936), p. 122.
[2] *Tables of the Incomplete Beta-function*, edited by Karl Pearson, (Office of *Biometrika*, University College, London), 1934, p. 494.
[3] Augustus Wadsworth and L. W. Hyman, *Jour. Immunol.*, Vol. 35 (1938), p. 55.
[4] W. R. Thompson, *Biometrika*, Vol. 25 (1933), p. 285.
[5] W. R. Thompson, *American Journal of Mathematics*, Vol. 57 (1935), p. 450.
[6] H. Hotelling and M. R. Pabst, *Annals of Mathematical Statistics*, Vol. 7 (1936), p. 29.
[7] Milton Friedman, *Jour. Amer. Stat. Assoc.*, Vol. 32 (1937), p. 675.

THE COMPUTATION OF MOMENTS WITH THE USE OF CUMULATIVE TOTALS

By Paul S. Dwyer

1. **Introduction.** Various authors have shown how the moments of a frequency distribution may be computed from cumulated frequencies.[1] In order to make clear to the reader the type of technique under discussion there is presented an illustration which is, essentially, that used by Hardy, [2, p. 59]. The value $\Sigma f_x = 729$ is the last entry in column 4.

We use C_1^1 to denote the entry in column 4 which is opposite the smallest variate (or class mark if the distribution is grouped). Similarly C_2^1 is the entry above C_1^1, and C_1^2 the entry to the right of C_1^1, etc. In this notation the diagonal entries, the ones underscored in Table I, are C_1^1, C_2^2, C_3^3, C_4^4, C_5^5.

The moments[2] about the smallest variate can be expressed in terms of the cumulations of Table I in different ways. One method utilizes the diagonal entries and the differences of zero. Thus

$$\sum_0^6 xf_x = C_2^2 = 2916; \qquad \sum_0^6 x^2f_x = C_2^2 + 2C_3^3 = 12636;$$

$$\sum_0^6 x^3f_x = C_2^2 + 6C_3^3 + 6C_4^4 = 57996;$$

$$\sum_0^6 x^4f_x = C_2^2 + 14C_3^3 + 36C_4^4 + 24C_5^5 = 278316, \text{ etc.}$$

A second method utilizes the entries in the next to the last row and the differences of zero. Thus

$$\sum_0^6 xf_x = C_2^2 = 2916; \qquad \sum_0^6 x^2f_x = -C_2^2 + 2C_2^3 = 12636;$$

$$\sum_0^6 x^3f_x = C_2^2 - 6C_2^3 + 6C_2 = 57996;$$

$$\sum_0^6 x^4f_x = -C_2^2 + 14C_2^3 - 36C_2^4 + 24C_2^5 = 278316, \text{ etc.}$$

---

[1] The reader is referred to reference [1] . . . [15], at end of paper.

[2] It is to be noted that we are not talking about moments *per unit frequency*. We are using the term in the sense used for example by Whittaker and Robinson. See [20, p. 18].

A third method, which seems to have escaped previous attention, involves columnar entries and multipliers whose determination and properties are a chief concern of this paper. Thus

$$\sum_0^6 xf_x = C_2^2 = 2916; \qquad \sum_0^6 x^2 f_x = C_2^3 + C_3^3 = 12636;$$

$$\sum_0^6 x^3 f_x = C_2^4 + 4C_3^4 + C_4^4 = 57996;$$

$$\sum_0^6 x^4 f_x = C_2^5 + 11C_3^5 + 11C_4^5 + C_5^5 = 278316, \text{ etc.}$$

It is possible also to obtain formulas when the cumulations are made from the smallest variate to the largest variate and, indeed, the whole theory of the present paper could be duplicated with such a theory of cumulation.

TABLE I

Successive Frequency Cumulations

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X	x	F_x	C^1	C^2	C^3	C^4	C^5
$a+6$	6	64	64	64	64	64	64
$a+5$	5	192	256	320	384	448	512
$a+4$	4	240	496	816	1200	1648	2160
$a+3$	3	160	656	1472	2672	4320	6480
$a+2$	2	60	716	2188	4860	9180	15660
$a+1$	1	12	728	2916	7776	16956	32616
a	0	1	729	3645	11421	28377	60993

It is possible to obtain the columnar formulas from the well known diagonal formulas. From the construction of Table I it is clear that

$$C_i^j = C_{i+1}^j + C_i^{j-1} \tag{1}$$

so that

$$\begin{aligned} &C_2^2 = C_2^2; C_2^2 + 2C_3^3 = C_2^3 + C_3^3; \qquad C_2^2 + 6C_3^3 + 6C_4^4 = C_2^4 + 4C_3^4 + C_4^4; \\ &\qquad C_2^2 + 14C_3^3 + 36C_4^4 + 24C_5^5 = C_2^5 + 11C_3^5 + 11C_4^5 + C_5^5. \end{aligned} \tag{2}$$

Formula (1) can be used similarly in deriving columnar formulas from row formulas, diagonal formulas from row formulas, etc.

The columnar method is here recommended as a useful substitute for the usual elementary method of computing moments. The many multiplications involved in the usual process are replaced by continued addition. The chief

disadvantage of the method is the continual recording, although this obstacle is surmounted with an adding machine equipped with a recording tape. The resulting moments are easily checked with an adaptation of Charlier's check, as is shown in section 8, and methods are given by which the multipliers are easily obtained. The method is also well adapted to the use of Hollerith machines.

The introduction of such columnar multipliers tends to give a different emphasis to the cumulative totals technique. The use of diagonal entries led logically to an emphasis upon factorial moments, while the columnar method tends to emphasize the more familiar power moments. The primary application here indicated is not to elaborate and specialized techniques, but rather to the simple, though often tedious, problem of the computation of power moments.

The aims of this paper are then:

(1) To show how moments may be computed from the columnar values of the successive cumulations,

(2) To discover the properties of the columnar multipliers,

(3) To present a general theory for computation of moments using cumulative totals.

2. The Basic Cumulative Theorem. The use of (1) is not satisfactory in getting precise formulas for the columnar multipliers so we derive the columnar cumulative theory directly from first principles. We first prove

THEOREM I. *Let x be any real number and let u_x be a real function of x which is 0 when $x < a$ and when $x > a + k$ and which is not infinite for $x = a, a + 1, a + 2, \cdots, a + k$. Let v_x be a real function of x and $\underline{v}_x$, called range v_x, a function such that $\underline{v}_x = v_x$ when $x = a, a + 1, \cdots, a + k$ and $\underline{v}_x = 0$ at all points outside the range a to $a + k$. If $\sum_x^{a+k} u_i$ is indicated by Cu_x and $\underline{v}_x - v_{x-1}$ by ∇v_x, $\underline{v}_x - \underline{v}_{x-1}$ by $\nabla \underline{v}_x$ then*

$$\sum_a^{a+k} u_x v_x = \sum_a^{a+k} u_x \underline{v}_x = \sum_a^{a+k} Cu_x \nabla \underline{v}_x. \tag{3}$$

The values u_x, v_x, Cu_x, $\nabla \underline{v}_x$ are presented in Table II.

The theorem is proved by forming

$$\sum_a^{a+k} Cu_x \nabla \underline{v}_x \equiv u_{a+k} v_{a+k} + \cdots + u_{a+i} v_{a+i} + \cdots + u_a v_a$$

$$\equiv \sum_a^{a+k} u_x v_x \equiv \sum_a^{a+k} u_x \underline{v}_x.$$

Theorem I can also be written as

$$\sum_0^k u_{a+x} v_{a+x} = \sum_0^k Cu_{a+x} \nabla \underline{v}_{a+x} \tag{4}$$

3. **The Successive Cumulation Theorem.**

THEOREM II. *If* $C^2u_x = C[Cu_x]$ *and* $\nabla^2\underline{v}_x = \nabla(\nabla\underline{v}_x)$, *etc., then*

$$\sum_a^{a+k} u_x v_x = \sum_a^{a+k} u_x \underline{v}_x = \sum_a^{a+k} C^{s+1}u_x\nabla^{s+1}\underline{v}_x .$$

This theorem follows readily from Theorem I. If

$$U_x = Cu_x \quad \text{and} \quad V_x = \nabla v_x , \qquad \text{then}$$

$$\sum_a^{a+k} u_x v_x = \sum_a^{a+k} u_x \underline{v}_x = \sum_a^{a+k} U_x \underline{V}_x = \sum_a^{a+k} CU_x \nabla \underline{V}_x = \sum_a^{a+k} C^2 u_x \nabla^2 \underline{v}_x .$$

This process can be extended as many times as desired so that

$$\sum_a^{a+k} u_x v_x = \sum_a^{a+k} u_x \underline{v}_x = \sum_a^{a+k} C^{s+1}u_x\nabla^{s+1}\underline{v}_x . \tag{5}$$

TABLE II

Values of x, u_x, v_x, Cu_x, *and* $\nabla\underline{v}_x$.

x	u_x	v_x	Cu_x	$\nabla\underline{v}_x$
$a+k$	u_{a+k}	v_{a+k}	u_{a+k}	$\underline{v}_{a+k} - \underline{v}_{a+k-1}$
$a+k-1$	u_{a+k-1}	v_{a+k-1}	$u_{a+k} + u_{a+k-1}$	$\underline{v}_{a+k-1} - \underline{v}_{a+k-2}$
....	....	....	....	....
$a+i$	u_{a+i}	v_{a+i}	$u_{a+k} + \cdots + u_{a+i}$	$\underline{v}_{a+i} - \underline{v}_{a+i-1}$
....	....	....	....	....
$a+1$	u_{a+1}	v_{a+1}	$u_{a+k} + \cdots + u_{a+i} + \cdots + u_{a+1}$	$\underline{v}_{a+1} - \underline{v}_a$
a	u_a	v_a	$u_{a+k} + \cdots + u_{a+i} + \cdots + u_{a+1} + u_a$	$\underline{v}_a$

This can also be written as

$$\sum_0^k u_{a+x}v_{a+x} = \sum_0^k u_{a+x}\underline{v}_{a+x} = \sum_0^k C^{s+1}u_{a+x}\nabla^{s+1}\underline{v}_{a+x} . \tag{6}$$

In order to determine the values $\nabla^{s+1}\underline{v}_{a+x}$, $0 \leq x \leq k$, we note that

$$\nabla^{s+1}v_{a+x} = \sum_0^s (-1)^t \binom{s+1}{t} v_{a+x-t} , \tag{7}$$

so that

$$\nabla^{s+1}\underline{v}_{a+x} = \sum_0^s (-1)^t \binom{s+1}{t} \underline{v}_{a+x-t} . \tag{8}$$

We also know that, $x \leq k$

$$
(9) \qquad \left.\begin{aligned} \underline{v}_{a+x-t} &= v_{a+x-t} && \text{when } t \leq x \\ \underline{v}_{a+x-t} &= 0 && \text{when } t > x \end{aligned}\right\}
$$

so that

$$
(10) \qquad \nabla^{s+1}\underline{v}_{a+x} = \sum_0^x (-1)^t \binom{s+1}{t} v_{a+x-t}, \qquad 0 \leq x \leq s
$$

$$
(11) \qquad \nabla^{s+1}\underline{v}_{a+x} = \sum_0^s (-1)^t \binom{s+1}{t} v_{a+x-t} = \nabla^{s+1} v_{a+x}, \qquad s < x \leq k.
$$

The formula (6) can then be written

$$
(12) \quad \sum_a^{a+k} u_x v_x = \sum_0^k u_{a+x} v_{a+x} = \sum_0^s C^{s+1} u_{a+x} \nabla^{s+1} \underline{v}_{a+x} + \sum_{s+1}^k C^{s+1} u_{a+x} \nabla^{s+1} v_{a+x}.
$$

4. **Moments from the Cumulated Frequencies.** If $u_{a+x} = f_{a+x}$ and $v_{a+x} = (a + x)^s$, then (6) gives

$$
(13) \qquad \sum_0^k (a + x)^s f_{a+x} = \sum_0^k C^{s+1} f_{a+x} \nabla^{s+1} \underline{(a + x)}^s.
$$

A more useful formula, obtained from (12), is

$$
(14) \qquad \sum_0^k (a + x)^s f_{a+x} = \sum_0^s C^{s+1} f_{a+x} \nabla^{s+1} \underline{(a + x)}^s,
$$

since $\nabla^{s+1} (a + x)^s = 0$. We have then

THEOREM III. *The values of the s-th moments can be obtained from the last $s + 1$ entries of the $(s + 1)$st cumulation of the frequencies. The multipliers are the values*

$$
(15) \qquad \nabla^{s+1} \underline{(a + x)}^s = \sum_0^x (-1)^t \binom{s+1}{t} (a + x - t)^s.
$$

Cor. 1. When $a = 0$, i.e., when the moments are measured about the smallest variate, the multipliers are

$$
(16) \qquad \nabla^{s+1} \underline{x}^s = \sum_0^x (-1)^t \binom{s+1}{t} (x - t)^s.
$$

Cor. 2. When $a = 1$, the multipliers are

$$
(17) \qquad \nabla^{s+1} \underline{(1 + x)}^s = \sum_0^x (-1)^t \binom{s+1}{t} (1 + x - t)^s.
$$

Cor. 3. If the moments are measured about a fixed value, p, then the new smallest variate is $a - p = a'$ and the multipliers are $\nabla^{s+1} \underline{(a' + x)^s}$.

Cor. 4. If p is the mean, m, then $a' = a - m$. If in addition $a = 0$, then $a' = -m$ and the multipliers giving moments about the mean are $\nabla^{s+1}\,\underline{(x - m)}^s$. Now

$$m = \frac{\sum_0^k xf_x}{\sum_0^k f_x} = \frac{C_1^2\nabla^2\underline{0} + C_2^2\nabla^2\underline{1}}{C_1^1} = \frac{C_2^2}{C_1^1}.$$

It follows that the multipliers giving the moments about the mean are

$$\nabla^{s+1}\underline{\left(x - \frac{C_2^2}{C_1^1}\right)}^s. \tag{18}$$

It is to be noted that the moments about different points are obtained by applying different multipliers to the same cumulated frequencies.

5. **Values of the multipliers.** The values of the multipliers may be computed from (15). Thus $\nabla^3\underline{(a + 1)}^2 = (a + 1)^2 - 3a^2 = -2a^2 + 2a + 1$. This becomes $2ab + 1$ when $1 - a$ is set equal to b. Values of the multipliers for the most common values of s and x are presented in Table III.

TABLE III

Values of $\nabla^{s+1}\,\underline{(a + x)}^s$

x \ s	0	1	2	3	4
4					b^4
3				b^3	$4b^3a + 6b^2 + 4b + 1$
2			b^2	$3b^2a + 3b + 1$	$6a^2b^2 + 12ab + 11$
1		b	$2ab + 1$	$3a^2b + 3a + 1$	$4a^3b + 6a^2 + 4a + 1$
0	1	a	a^2	a^3	a^4

When $a = 0, b = 1$ and the multipliers are 1; 0, 1, 1; 0, 1, 4, 1; 0, 1, 11, 11, 1; etc. as indicated in section 1. When $a = 1, b = 0$ and the multipliers are 1; 1, 0; 1, 4, 1, 0; 1, 11, 11, 1, 0; etc. When the moments are measured about a fixed point, p, it is only necessary to compute $a' = a - p$ and to use a' for a and $b' = 1 - a'$ for b in Table III.

We illustrate the use of the multipliers by application to the problem of Table I. The moments about the smallest variate are computed in section 1. The moments, when $a = 1$ are $\sum_0^6 (x + 1)f_x = C_1^2 = 3654$; $\sum_0^6 (x + 1)^2f_x =$

$C_1^3 + C_2^3 = 19197; \sum_0^6 (x + 1)^3 f_x = C_2^4 + 4C_3^4 + C_4^4 = 105381; \sum_0^6 (x + 1)^4 f_x = C_2^5 + 11C_3^5 + 11C_4^5 + C_5^5 = 598509.$

The moments about the mean are found by forming $\dfrac{C_2^2}{C_1^1} = \dfrac{2916}{729} = 4$. Then $a = -4$ and the multipliers are 1; -4, 5; 16, -39, 25; -64, 229, -284, 125; 256, -1199, 2171, -1829, 625; etc. so that $\sum_0^6 \bar{x} f_x = 0$; $\sum_0^6 \bar{x}^2 f_x = 972$; $\sum_0^6 \bar{x}^3 f_x = -324$; $\sum_0^6 \bar{x}^4 f_x = 3564$.

Since the values of $\nabla^{s+1}(x - C_2^2/C_1^1)^s$ are expressible in terms of C_1^1 and C_2^2, it follows that the values of $\sum_0^k \bar{x}^s f_x$ are expressible in terms of cumulations. For example a formula for the second moment about the mean, which is essentially one given by Whittaker and Robinson [7, p. 193] is

$$\sum_a^{a+k} \bar{x}^2 f_x = C_2^2 + 2C_3^3 - \frac{(C_2^2)^2}{C_1^1}. \tag{19}$$

However the general method described above, supplemented with the techniques of succeeding sections, is preferred to the development and use of such formulas.

6. **Recursion Property of the Multipliers.** It is not readily apparent from Table III how the multipliers of the $(s + 1)$-th cumulations can be obtained from the multipliers of the s-th cumulations. It is possible to establish a recursion formula which is useful for this purpose. Now, $a \leq x \leq s$,

$$\nabla^{s+1}\underline{(a + x)}^s = (a + x)^s + \sum_1^x (-1)^t \binom{s+1}{t} (a + x - t)^s$$

$$(a + x)\nabla^s\underline{(a + x)}^{s-1} = (a + x)^s + \sum_1^x (-1)^t \binom{s}{t} (a + x - t)^{s-1}(a + x)$$

$$(s + 1 - a - x)\nabla^s\underline{(a + x - 1)}^{s-1}$$
$$= \sum_1^x (-1)^{t-1} \binom{s}{t-1} (a + x - t)^{s-1}(s + 1 - a - x)$$

and since

$$\binom{s}{t}(a + x) - \binom{s}{t-1}(s + 1 - a - x) = \binom{s+1}{t}(a + x - t)$$

it follows that

$$\Delta^{s+1}\underline{(a + x)}^s = (a + x)\nabla^s\underline{(a + x)}^{s-1} + (s + 1 - a - x)\nabla^s\underline{(a + x - 1)}^{s-1}. \tag{20}$$

When $a = 0$ we have

$$\nabla^{s+1}\underline{x}^s = x\nabla^s\underline{x}^{s-1} + (s + 1 - x)\nabla^s(\underline{x - 1})^{s-1}. \tag{21}$$

Formulas (20) and (21), though somewhat formidable in appearance, are easy to apply. Thus $\nabla^3(\underline{a + 2})^2 = (a + 2)\nabla^2(\underline{a + 2}) + (1 - a)\nabla^2(\underline{a + 1})$. The recursion formula is especially useful in building up tables of multipliers. The following form is recommended:

As successive columnar headings use the values a, $a + 1$, $a + 2$, etc. and as successive row headings use $1 - a$, $2 - a$, $3 - a$, etc. Then $\nabla\underline{a}^0 = 1$ is placed in the upper left cell, $\nabla^2\underline{a}$ directly below $\nabla\underline{a}^0$, $\nabla^2\underline{a + 1}$ to the right of $\nabla^2\underline{a}^0$, etc. The values of $\nabla^3(\underline{a + x})^2$ are placed in the next diagonal, etc. If this process is continued the entry $\nabla^{s+1}(\underline{a + x})^s$ will have the entry $\nabla^s(\underline{a + x})^{s-1}$ directly above it and the entry $\nabla^s(\underline{a + x - 1})^{s-1}$ on its left. Also the columnar heading is $a + x$ and the row heading $s + 1 - a - x$ so that any entry is obtained by adding the product of the entry above it and the columnar heading to the product of the entry to the left and the row heading. The values of $\nabla^{s+1}\underline{x}^s$ are obtained by placing $a = 0$. They are presented, in Table IV, through $s = 8$.

TABLE IV

Values of $\nabla^{s+1}\underline{x}^s$

$s+1-x$ \ x	0	1	2	3	4	5	6	7	8
1	1	1	1	1	1	1	1	1	1
2	0	1	4	11	26	57	120	247	
3	0	1	11	66	302	1191	4293		
4	0	1	26	302	2416	15619			
5	0	1	57	1191	15619				
6	0	1	120	4293					
7	0	1	247						
8	0	1							

The table is easily extended to higher values of s. If a table of values of ∇^{s+1} $(\underline{x + 1})^s$ is constructed, it will be found to be like Table IV with columns and rows interchanged. Hence the values of $\nabla^{s+1}(\underline{x + 1})^s$ are obtained from Table

IV by reading the multipliers down the diagonal. Thus the values $\nabla^3\underline{(x+1)}^2$ are 1, 4, 1, 0, etc.

The ease with which the multipliers may be computed is illustrated with $a = -4$. In this case we have

TABLE V

Values of $\nabla^{s+1}\,\underline{(x+a)}^s$ *with* $a = -4$

$s+1-a-x$ \ $a+x$	-4	-3	-2	-1	0
5	1	5	25	125	625
6	-4	-39	-284	-1829	
7	16	229	2171		
8	-64	-1199			
9	256				

These values agree with those computed more laboriously in section 5.

7. **Value of** $\sum_0^k \nabla^{s+1}\underline{(x+a)}^s$. It is to be noted in Tables III, IV, V that the sum of the entries in the diagonal having $s+1$ terms is $s!$ This is generally true and results from the fact that

$$\sum_0^k \nabla^{s+1}\underline{(x+a)}^s = \sum_0^s \nabla^{s+1}\underline{(x+a)}^s = s! \tag{22}$$

In obtaining the values of $\sum_0^k \nabla^{s+2}\underline{(x+a)}^{s+1}$ from the value of $\sum_0^k \nabla^{s+1}\underline{(x+a)}^s$ it is noted that $\nabla^{s+1}\underline{(x+a)}^s$ is used but twice. Once it is multiplied by $a+x$ and once by $s+1-a-x$ so that the net result is a multiplication by $s+1$. It follows that $\sum_0^k \nabla^{s+2}\underline{(x+a)}^{s+1} = (s+1)\sum_0^k \nabla^{s+1}\underline{(x+a)}^s$ and since $\sum_0^k \nabla^2\underline{(x+1)} = 1$, $\sum_0^k \nabla^3\underline{(x+a)}^2 = 2!$ so that in general $\sum_0^k \nabla^{s+1}\underline{(x+a)}^s = s!$

This property is useful in checking the values of the computed multipliers.

8. **The adaptation of the Charlier check.** An adaptation of the Charlier check serves as an excellent check for the computed moments. It is recalled that the Charlier check gives

$$\sum_a^{a+k} (x+1)^s f_x = \sum_{t=0}^{s} \binom{s}{t} \sum_{x=a}^{a+k} x^{s-t} f_x. \tag{23}$$

The components of the right hand member are computed by cumulative totals as indicated above. The left hand member is obtained by applying different multipliers to the same cumulated frequencies. Thus $\sum_{a}^{a+k} (x+1)^s f_x = \sum_{0}^{k} (x + a + 1)^s f_{x+a}$ and the multipliers of the cumulated frequencies are $\nabla^{s+1}\underline{(x + a')^s}$ where $a' = a + 1$. If $a = 0$ the Charlier check multipliers are the values $\nabla^{s+1}\underline{(x+1)^s}$ which can be read from Table IV. For example $\sum_{0}^{6} (x+1)^4 f_x = C_1^5 + 11C_2^5 + 11C_3^5 + C_4^5 = 598509$ and this checks with $\sum_{0}^{6} x^4 f_x + 4\sum_{0}^{6} x^3 f_x + 6\sum_{0}^{6} x^3 f_x + 4\sum_{0}^{6} x f_x + \sum_{0}^{6} f_x$.

9. **Application to factorial moments**. When $u_x = f_x$, $v_x = x^{(s)} = x(x-1)(x-2)\cdots(x-s+1)$

$$\sum_{0}^{k} x^{(s)} f_x = \sum_{0}^{k} C^{s+1} f_x \nabla^{s+1} \underline{x}^{(s)}$$

and since $\nabla^{s+1}\underline{x}^{(s)}$ is 0 when $s < x \leq k$, is $s!$ when $s = x$, is 0 when $0 \leq x < s$,

$$\sum_{0}^{k} x^{(s)} f_x = \sum_{s}^{k} x^{(s)} f_x = s!\, C_{s+1}^{s+1}. \tag{24}$$

It follows that the underscored terms of Table I, when multiplied by $s!$, give the factorial moments. Factorial moments, first used by Sheppard [4], have since come into prominence largely because of this ease of computation.

The coefficients of $(a + b)^x$ are $1, x, \frac{x(x-1)}{2!}, \cdots, \frac{x^{(s)}}{s!}, \cdots$. If we define the binomial moment by $B_s = \sum_{0}^{k} \frac{x^{(s)}}{s!} f_x$ [6, p. 278] then $B_s = \frac{1}{s!}\sum_{0}^{k} x^{(s)} f_x = C_{s+1}^{s+1}$.

It is also possible to show that the entries under the main diagonal are binomial moments. In Table I, for example, we let $a = 1$ and add the additional row $a = 0$ with 0 frequency. Then $C_1^1 = 729$, $C_2^1 = 729$, $C_1^2 = 729 + 3645 = 4374$, etc. The new diagonal terms are directly under the old diagonal terms and give $B_{s,1} = \sum_{1}^{7} x^{(s)} f_x = \sum_{0}^{6} (x+1)^{(s)} f_x$. In general the terms $B_{s,l}$ are given l rows below the terms B_s and the factorial moments are $s!\, B_{s,l}$. Then

$$F_{s,l} = s!\, C_{s+1-l}^{s+1}. \tag{25}$$

For example in the problem of Table I, $F_{4,3} = \sum_{3}^{9} x^{(4)} f_x = 4! C_2^5 = 782{,}784$. The method is especially adapted to the use of Hollerith machines, for positive integral values of l, since it is only necessary to have the machine continue its cumulation.

10. **The cumulations of** xf_x. It is possible to use the cumulations of xf_x in securing the values of the moments. Now

$$\sum_{a}^{a+k} x^{s+1} f_x = \sum_{0}^{k} (x+a)^{s+1} f_{x+a} = \sum_{0}^{k} (x+a) f_{x+a} (x+a)^{s} \tag{26}$$

$$= \sum_{0}^{s} C^{s+1} (x+a) f_{x+a} \nabla^{s+1} \underline{(x+a)}^{s}.$$

When $a = 0$, (26) becomes

$$\sum_{0}^{k} x^{s+1} f_x = \sum_{0}^{s} C^{s+1} x f_x \nabla^{s+1} \underline{x}^{s}. \tag{27}$$

We compute the cumulations of xf for the problem of Table I. These are given in Table VI.

TABLE VI

Cumulations of xf_x

x	f_x	xf_x	C^1	C^2	C^3	C^4
6	64	384	384	384	384	384
5	192	960	1344	1728	2112	2496
4	240	960	2304	4032	6144	8640
3	160	480	2784	6816	12960	21600
2	60	120	2904	9720	22680	44280
1	12	12	2916	12636	35316	79596
0	1	1	2916	15552	50868	130464

so that

$$\sum_{0}^{6} xf_x = 2916; \qquad \sum_{0}^{6} x^2 f_x = 12636; \qquad \sum_{0}^{6} x^3 f_x = 35316 + 22680 = 57996;$$

$$\sum_{0}^{6} x^4 f_x = 79596 + 4(44280) + 21600 = 278316.$$

In getting moments about the mean from the cumulations of xf_x, the following method is recommended.

$$\sum_{0}^{k} \bar{x}^{s+1} f_x = \sum_{0}^{k} \bar{x}^{s} (x - m) f_x = \sum_{0}^{k} \bar{x}^{s} x f_x - m \sum_{0}^{k} \bar{x}^{s} f_x. \tag{28}$$

and

$$\sum_{0}^{k} \bar{x}^{s} x f_x = \sum_{0}^{k} C^{s+1} (x f_x) \nabla^{s+1} (x - m)^{s}. \tag{29}$$

When $s = 1$, (28) gives $\sum_{0}^{k} \bar{x}^2 f_x = \sum_{0}^{k} \bar{x} x f_x - m \sum_{0}^{k} \bar{x} f_x$ and

(30)
$$\sum_0^k \bar{x}^2 f_x = \sum_0^k \bar{x} x f_x .$$

In the illustrative problem $a = -4$ so that

$$\sum_0^6 \bar{x} x f_x = -4(15552) + 5(12636) = 972$$

$$\sum_0^6 \bar{x}^2 x f_x = 16(50868) - 39(35316) + 25(22680) = 3564$$

$$\sum_0^6 \bar{x}^3 x f_x = 2268$$

and

$$\sum_0^6 \bar{x}^2 f_x = 972; \qquad \sum \bar{x}^3 f_x = 3564 - 4(972) = -324; \qquad \sum \bar{x}^4 f_x = 3564.$$

Formula (30) is of note since it permits the determination of $\sum_0^k \bar{x}^2 f_x$ directly from the cumulations of $x f_x$.

The factorial moments are also related to the cumulations of $x f_x$. Thus

(31)
$$\sum_0^k x^{(s)} f_x = \sum_0^k (x-1)^{(s-1)} x f_x = \sum_0^k C^s(x f_x) \nabla^s \underline{(x-1)^{(s-1)}}$$

which results in $\sum_0^k x^{(s)} f_x = (s-1)! C_s^s(x f_x)$.

It follows that

$$C_s^s(x f_x) = s C_{s+1}^{s+1}(f_x).$$

For example, the underscored terms of Table VI are respectively 1, 2, 3, 4 times underscored terms of Table I.

In general the cumulations of $x f_x$, rather than of f_x, are recommended since $C(x f_x)$ can be computed and recorded almost as quickly as $C(f_x)$, since one less cumulation is needed to obtain a specific moment, and since the multipliers needed to get a specific moment are smaller. A technique based on the cumulations of $x f_x$ is especially adapted to the use of Hollerith machines. Let us take x_x to represent the sum of the x's for all items in the distribution having the same value of x. Then $x f_x = x_x$ and we have

(32)
$$\sum_a^{a+k} x^s f_x = \sum_a^{a+k} x^{s-1} x_x = \sum_a^{a+k} C^s(x_x) \nabla^s \underline{(x^{s-1})}.$$

If the cards are sorted for x and the tabulator is wired to print cumulative totals each time x changes, the recording tape gives the successive values of $C(x_x)$. (Care must be taken that there are no blank values of x.)

If a summary punch is available, these cumulations are punched on cards as

they are cumulated and these summary cards are used in getting higher cumulations.

If no summary punch is available, it is possible to obtain $\sum x^2 f_x$ by the application of Theorem I. Thus

$$\sum_a^{a+k} x^2 f_x = \sum_a^{a+k} x x_x = \sum_a^{a+k} C(x_x) \nabla(\underline{x}),$$

and since $\nabla(\underline{x}) = a$ when $x = a$ and $\nabla(\underline{x}) = 1$ when $x > a$, it follows that $\sum_a^{a+k} x^2 f_x$ can be obtained by adding the entries above the last and then adding the last entry multiplied by a. This is essentially the Mendenhall-Warren-Hollerith method of getting $\sum x^2 f_x$ [9, p. 27].

In case $a = 0$ the technique amounts simply to adding all the entries above the bottom one.

The value $\sum x^3 f_x$ can be obtained similarly from the first order cumulations. Thus

$$\sum_a^{a+k} x^3 f_x = \sum_a^{a+k} x^2 x_x = \sum_a^{a+k} C(x_x) \nabla(\underline{x^2}) \tag{33}$$

and since $\nabla(\underline{x^2}) = a^2$ when $x = a$, $\nabla(\underline{x^2}) = 2x - 1$ when $x > a$, it follows that

$$\sum_a^{a+k} x^3 f_x = a^2 C_1^1(x_x) + \sum_{a+1}^{a+k} C(x_x)(2x - 1). \tag{34}$$

When $a = 0$, (34) becomes

$$\sum_0^k x^3 f_x = \sum_1^k C(x_x)(2x - 1) \tag{35}$$

so that the multipliers are the successive odd integers. Thus from the first order cumulations of Table VI we have

$$\sum_0^6 x f_x = 2916; \qquad \sum_0^6 x^2 f_x = 12636; \qquad \sum_0^6 x^3 f_x = 57996.$$

The cumulative method can also be applied to the method of digiting [17, p. 425].

It is also possible to obtain the moments from the cumulations of $x^2 f_x$, $x^3 f_x$, etc., since

$$\sum_a^{a+k} x^{s+2} f_x = \sum_a^{a+k} x^s x^2 f_x = \sum_a^{a+k} C^{s+1}(x^2 f_x) \nabla^{s+1}(\underline{x^s})$$

$$\sum_a^{a+k} x^{s+3} f_x = \sum_a^{a+k} x^s x^3 f_x = \sum_a^{a+k} C^{s+1}(x^3 f_x) \nabla^{s+1}(\underline{x^s})$$

but the cumulations of $x f_x$ are preferable for most purposes. The Charlier check works in all cases. It should be noted that the indicated Hollerith technique

demands only the customary tabulator and not the expensive, time consuming, card punching, multiplier, [16].

11. **Product Moments. Correlation.** It is possible to apply the cumulative technique in getting product moments involving two variables. If we let y_x be the sum of all the values of y having the same value of x, then

$$\sum x^s y f_{xy} = \sum_a^{a+k} y_x x^s = \sum_a^{a+s} C^{s+1}(y_x)\nabla^{s+1}(\underline{x^s}) \tag{36}$$

so that the multipliers are the same as those previously used. When Hollerith machines are used, it is only necessary to sort the cards for x and to wire the machine to give cumulations on variables x, y, z, etc. If the machine is adjusted to take totals with each change in x, the tape records simultaneously the values of $C(x_x)$, $C(y_x)$, $C(z_x)$, etc. With a summary punch it is possible to form successive cumulations easily. The values Σx^{s+1}, $\Sigma x^s y$, $\Sigma x^s z$, etc. are then obtained by applying the multipliers. When $s = 1$, (36) becomes

$$\sum xy f_{xy} = \sum_a^{a+k} C^2(y_x)\nabla^2(\underline{x}) \tag{37}$$

so that the multipliers are a, $1 - a$, 0, 0, etc. When $a = 0$, the multipliers are 0, 1, 0, 0, etc. and when $a = 1$, they are 1, 0, 0, etc.

When no summary punch is available, it is necessary to obtain the values of the moments from the first order cumulations. Using Theorem I

$$\sum xy f_{xy} = \sum_a^{a+k} C(y_x)\nabla(\underline{x}) = aC_1^1(y_x) + \sum_{a+1}^{a+s} C(y_x). \tag{38}$$

This formula serves as the basis of the Mendenhall-Warren-Hollerith Correlation Method, [9, p. 27].

It can be shown in similar fashion that

$$\sum x^2 y f_{xy} = a^2 C_1^1 + \sum_{a+1}^{a+s} C(y_x)(2x - 1) \tag{39}$$

and when $a = 0$

$$\sum x^2 y f_{xy} = \sum_1^s C(y_x)(2x - 1). \tag{40}$$

The method is also adapted to the common problem of finding correlation coefficients from grouped data when Hollerith machines are not available and this method is recommended for the determination of these coefficients.

An illustration is presented in Table VII which shows the correlation existing between college first semester average, X, and preparatory school average, Y, for 1126 students entering the College of Literature, Science and the Arts of the University of Michigan in 1928. The coded values of X and Y are indicated by x and y and are positive integers beginning with 0. The coded values are given

in descending order beginning with the upper left hand corner of the chart. The values of the cumulations are placed at the right hand side and at the bottom of the chart.

TABLE VII

Correlation with cumulative totals

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
X / Y			4.00	3.99 3.50–	3.49 3.00–	2.99 2.50–	2.49 2.00–	1.99 1.50–	1.49 1.00–	.99 .50–	.49 .00–		
	x / y		8	7	6	5	4	3	2	1	0		
		f_x / f_y	13	50	107	220	341	179	121	60	35	Cx_y	Cy_y
4.00	6	18	5	2	5	5	1					113	108
3.99 3.50–	5	106	2	19	29	27	20	7		1	1	673	638
3.49 3.00–	4	178	3	12	35	53	44	18	6	5	2	1503	1350
2.99 2.50–	3	270	3	10	20	55	103	33	27	11	8	2568	2160
2.49 2.00–	2	330		6	11	54	114	67	46	19	13	3714	2820
1.99 1.50–	1	173		1	5	19	45	44	34	18	7	4244	2993
1.49 1.00–	0	51			2	7	14	10	8	6	4	4399	2993
		Cy_x	61	259	661	1330	2194	2578	2809	2923	2993	12815	10069
		Cx_x	104	454	1096	2196	3560	4097	4339	4399	4399	20245	1126

The lower right hand corner has the entries

$$\left.\begin{array}{ccc} & \sum x & \sum y \\ \sum y & \sum xy & \sum y^2 \\ \sum x & \sum x^2 & \sum f = N \end{array}\right\}$$

where $\sum xy$, $\sum y$, and $\sum x$ are obtained by adding the cumulations in the columns or rows involved.

The values $C(y_y)$ are easily computed from columns (2) and (3). The values of $C(x_y)$ are computed by forming the cumulated product of the row frequency and x. The values are recorded when the products contributed by a given row have been computed. The values $C(y_x)$ and $C(x_x)$ are obtained similarly.

The value of r is easily obtained from the lower right hand entries. The value $A_{x,y} = N\Sigma xy - (\Sigma x)(\Sigma y)$ is obtained from diagonal entries, $A_{x,x} = N\Sigma x^2 - (\Sigma x)^2$

is obtained from entries in the last row, $A_{y,y} = N\Sigma y^2 - (\Sigma y)^2$ is obtained from the last column, and $r = \dfrac{A_{x,y}}{\sqrt{A_{x,x}A_{y,y}}}$ is easily computed. In the above problem $r = .441$.

The values M_x, M_y, σ_x, σ_y are also easily obtained from the lower right hand entries. The successive steps are indicated by the form

	Σx	Σy			M_y
Σy	Σxy	Σy^2			
Σx	Σx^2	N	$A_{x,x}$		
		$A_{y,y}$	$A_{x,y}$	$\sqrt{A_{y,y}}$	σ_y
			$\sqrt{A_{x,x}}$	$\sqrt{A_{x,x}A_{y,y}}$	
M_x			σ_x		r

Recent methods of applying cumulative totals theory to correlation are found in references [9], [14], [17], [18], [19].

The third order moments are obtained by multiplying the entries of $C(x_y)$, $C(y_y)$, $C(x_x)$, $C(y_x)$ by 1, 3, 5, etc. as indicated by (40). Thus $\Sigma x^3 f_x = 4399 + 3(4339) +$ etc. $= 102, 103$; $\Sigma x^2 y f_{xy} = 63121$; $\Sigma x y^2 f_{xy} = 46047$; $\Sigma y^3 f_y = 38{,}633$. It is hence possible to compute the skewness of each marginal distribution from Table VII. See also [18, p. 657].

12. **Conclusion**. This paper presents an outline of the computation of moments with the use of cumulative totals and columnar multipliers. Basic general theorems are derived and applications are made to one variable and two variable distributions both with and without punched card equipment. The formulas assume that the distance between successive variates (or class marks) is unity, but the reader should have no trouble in adapting the formulas to more general problems.

In the interest of brevity the development is limited to the descending cumulations. It is possible to parallel the development here by deriving formulas in terms of ascending cumulations. It is also possible to work out formulas showing relations between columnar, row, and diagonal multipliers. There are other applications such as to the evaluation of $\sum_1^K x^s$, which are of interest. It is possible also that applications may be found for the general theory of sections 2 and 3 which do not demand that v_x be a power function.

THE UNIVERSITY OF MICHIGAN.

REFERENCES

[1] G. F. LIPPS, "Die Theorie der Kollectivgegenstände," *Philosophische Studien* (Wundt Editor), Vol. 17, (1901) pp. 467-575.

[2] G. F. HARDY, *Theory of the Construction of Tables of Mortality*, pp. 59-62 and 124-128.

[3] W. P. ELDERTON, *Frequency Curves and Correlation*, pp. 19-23.

[4] W. F. SHEPPARD, "Factorial Moments in Terms of Sums or Differences," *Proc. of London Math. Society*, 2, Vol. 13, (1913) pp. 81-96.
Also, "Fitting Polynomials by the Method of Least Squares," ibid. pp. 97-108.

[5] J. STEFFENSEN, "Factorial Moments and Discontinuous Frequency Functions," *Skandinavisk Aktuarietidskrift*, 6, (1923) pp. 73-89.

[6] J. STEFFENSEN, *Interpolation*, pp. 93-104.

[7] E. T. WHITTAKER AND G. ROBINSON, *Calculus of Observations*, pp. 191-194.

[8] R. Frisch, "Sur le calcul numérique des moments ordinaires et des moments composés d'une distribution statistique," *Skandinavisk Aktuarietidskrift*, Vol. 10, (1927) pp. 81-91.

[9] R. M. MENDENHALL AND R. WARREN, "The Mendenhall-Warren-Hollerith Correlation Method," *Columbia University Statistical Bureau Document* No. 1.1929, Columbia University, New York, 43 pp.

[10] R. M. MENDENHALL AND R. WARREN, "Computing Statistical Coefficients from Punched Cards," *Jour. of Ed. Psy.*, Vol. 21, (1930) pp. 53-62.

[11] C. JORDAN, "Approximation and Graduation According to the Principle of Least Squares by Orthogonal Polynomials," *Annals of Math. Stat.*, 3, (1932) pp. 257-358.

[12] A. C. AITKIN, "On the Graduation of Data by the Orthogonal Polynomials of Least Squares," *Proc. of Roy. Soc. of Edin.*, Vol. 53, pp. 54-78.

[13] A. C. AITKIN, "On Fitting Polynomials to Weighted Data by Least Squares," *Proc. of Roy. Soc. of Edin.*, Vol. 54, (1933-34) pp. 1-11.

[14] CHEN-NAN LI, "Summation Method of Fitting Parabolic Curves and Calculating Linear and Curvilinear Coefficients on a Scatter Diagram," *Jour. of Am. Stat. Assn.*, 29, (1934) pp. 405-409.

[15] M. SASULY, *Trend Analysis of Statistics*, Chap. VIII. Also page 5.

[16] H. C. CARVER, "Uses of the Automatic Multiplying Punch"; *Punched Card Method in Colleges and Universities*, pp. 417-422.

[17] A. E. BRANDT, "Uses of the Progressive Digit Method"; *Punched Card Method in Colleges and Universities*, pp. 423-436.

[18] P. S. DWYER AND A. D. MEACHAM, "The Preparation of Correlation Tables on a Tabulator Equipped with Digit Selection," *Jour. Am. Stat. Assn.*, Vol. 32, (1937) pp. 654-662.

[19] W. N. DUROST AND H. M. WALKER, *Durost-Walker Correlation Chart* World Book Co. N. Y. (1938).

[20] H. L. RIETZ, *Mathematical Statistics*, 1927.

A NOTE ON THE DERIVATION OF FORMULAE FOR MULTIPLE AND PARTIAL CORRELATION*

By Louis Guttman

1. **Multiple Correlation.** Let the measurements of N individuals on each of the n variables $x_1, x_2, \cdots, x_k, \cdots, x_n$, be expressed as relative deviates; that is, such that

$$\Sigma x_k = 0, \qquad \Sigma x_k^2 = N, \qquad k = 1, 2, 3, \cdots, n,$$

where the summations extend over the N individuals.

If values of λ_k are determined so that

$$\Sigma(x_1 - \lambda_2 x_2 - \lambda_3 x_3 - \cdots - \lambda_n x_n)^2 \text{ is a minimum,}$$

and if we let

$$X_1 = \lambda_2 x_2 + \lambda_3 x_3 + \cdots + \lambda_n x_n, \tag{1}$$

then the multiple correlation coefficient, obtained from the regression of x_1 on the remaining $n - 1$ variables, is defined as

$$r_{1.234\cdots n} = r_{x_1 X_1}.$$

The square of the standard error of estimate of x_1 on the remaining $n - 1$ variables is defined as

$$\sigma^2_{1.234\cdots n} = \frac{1}{N}\Sigma(x_1 - X_1)^2.$$

The minimizing values for λ_k are obtained from the normal equations

$$\Sigma(x_1 - X_1)x_k = 0, \qquad k = 2, 3, \cdots, n. \tag{2}$$

which may be written in expanded notation as,

$$\begin{array}{l}
\lambda_2 + r_{23}\lambda_3 + r_{24}\lambda_4 + \cdots + r_{2n}\lambda_n = r_{12} \\
r_{32}\lambda_2 + \lambda_3 + r_{34}\lambda_4 + \cdots + r_{3n}\lambda_n = r_{13} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
r_{n2}\lambda_2 + r_{n3}\lambda_3 + r_{n4}\lambda_4 + \cdots + \lambda_n = r_{1n}
\end{array}$$

where $r_{jk} = \frac{1}{N}\Sigma x_j x_k = r_{kj}$, $r_{jj} = 1$.

---

* The notions involved in this demonstration are certainly well-known. However, the directness and simplicity of the derivations may lend some merit to their exhibition. The writer is indebted to Professor Dunham Jackson for useful advice.

From Cramer's rule it is seen that

$$\lambda_k = -\frac{R_{1k}}{R_{11}}, \quad \text{if} \quad k \neq 1, \ R_{11} \neq 0,$$

where R_{jk} is the cofactor of r_{jk} (or of r_{kj}) in the symmetric determinant

$$R = \begin{vmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & r_{jk} & \cdots \\ r_{n1} & r_{n2} & r_{n3} & \cdots & r_{nn} \end{vmatrix}.$$

Summing both sides of (1) over the N individuals shows that $\Sigma X_1 = 0$, so that the variance of X_1 is

$$\sigma_{X_1}^2 = \frac{1}{N} \Sigma X_1^2.$$

From (2), the residual $(x_1 - X_1)$ is orthogonal to each of the x_k except x_1; therefore the residual is orthogonal to any linear combination of these x_k and in particular to X_1; that is,

(3) $$\Sigma(x_1 - X_1)X_1 = 0,$$

or

$$\sigma_{X_1} r_{x_1 X_1} = \sigma_{X_1}^2$$

and therefore

(4) $$r_{x_1 X_1} = \sigma_{X_1}.$$

Multiplying both sides of (1) by $\frac{x_1}{N}$ and summing over the individuals, we get:

$$\begin{aligned} \sigma_{X_1} r_{x_1 X_1} &= r_{12}\lambda_2 + r_{13}\lambda_3 + \cdots + r_{1n}\lambda_n \\ &= -\frac{1}{R_{11}}(r_{12}R_{12} + r_{13}R_{13} + \cdots + r_{1n}R_{1n}) \\ &= 1 - \frac{R}{R_{11}}. \end{aligned}$$

From (4) then,

$$r_{1.234\cdots n}^2 = 1 - \frac{R}{R_{11}}.$$

It is clear that in general

$$r_{k.123\cdots,k-1,k+1,\cdots n}^2 = 1 - \frac{R}{R_{kk}}.$$

To find the standard error of estimate, expand

$$\frac{1}{N}\Sigma(x_1 - X_1)^2 = 1 - 2\sigma_{X_1} r_{x_1 X_1} + \sigma^2_{X_1}$$
$$= 1 - r^2_{x_1 X_1}$$
$$= \frac{R}{R_{11}}.$$

In general, when $\sigma_k = 1$,

(5) $$\sigma^2_{k.123\cdots,k-1,k+1,\cdots n} = \frac{R}{R_{kk}}.$$

2. **Partial Correlation.** If values of μ_k and ν_k are determined so that

$$\Sigma(x_1 - \mu_3 x_3 - \mu_4 x_4 - \cdots - \mu_n x_n)^2 \text{ is a minimum}$$

and $$\Sigma(x_2 - \nu_3 x_3 - \nu_4 x_4 - \cdots - \nu_n x_n)^2 \text{ is a minimum,}$$

and if we let

(6) $$Y_1 = \mu_3 x_3 + \mu_4 x_4 + \cdots + \mu_n x_n$$
$$Y_2 = \nu_3 x_3 + \nu_4 x_4 + \cdots + \nu_n x_n,$$

then the partial correlation coefficient between x_1 and x_2, holding the remaining $n - 2$ variables constant, is defined as

$$r_{12.34\cdots n} = r_{(x_1 - Y_1)(x_2 - Y_2)};$$

and since $\Sigma(x_k - Y_k) = 0$,

(7) $$r_{12.34\cdots n} = \frac{\frac{1}{N}\Sigma(x_1 - Y_1)(x_2 - Y_2)}{\sigma_{1.34\cdots n}\sigma_{2.34\cdots n}}.$$

Each μ_k is the negative of the ratio of the cofactor of r_{1k} to the cofactor of r_{11} in the determinant obtained by striking out the second row and the second column from R. We shall use the notation $R_{hi\text{-}jk}$ to mean the algebraic complement of the second order minor in R, whose complement is obtained by striking out row h and column i and then row j and column k. Then

$$\mu_k = \frac{R_{22\text{-}1k}}{R_{22\text{-}11}}.$$

By argument similar to that used in (3),

$$\Sigma(x_1 - Y_1)Y_2 = 0,$$

or

$$\Sigma x_1 Y_2 = \Sigma Y_1 Y_2.$$

Similarly,

$$\Sigma x_2 Y_1 = \Sigma Y_1 Y_2 .$$

Then the numerator of the right member of (7) becomes, after expanding and collecting terms,

$$r_{12} - \sigma_{Y_1} r_{x_2 Y_1} . \tag{8}$$

Multiplying both sides of (6) by $\frac{x_2}{N}$ and summing over the N individuals, we have,

$$\begin{aligned} \sigma_{Y_1} r_{x_2 Y_1} &= r_{23}\mu_3 + r_{24}\mu_4 + \cdots + r_{2n}\mu_n \\ &= \frac{1}{R_{22\text{-}11}} (r_{23} R_{22\text{-}13} + r_{24} R_{22\text{-}14} + \cdots + r_{2n} R_{22\text{-}1n}) \\ &= r_{12} + \frac{R_{12}}{R_{22\text{-}11}} . \end{aligned} \tag{9}$$

Analogous to (5), we have,

$$\sigma^2_{1.34\cdots n} = \frac{R_{22}}{R_{22\text{-}11}}, \qquad \sigma^2_{2.34\cdots n} = \frac{R_{11}}{R_{11\text{-}22}} . \tag{10}$$

From (8), (9), and (10) the right member of (7) becomes

$$\frac{-R_{12}}{\sqrt{R_{11} R_{22}}} .$$

It is seen that in general

$$r_{jk.12\cdots,j-1,j+1,\cdots,k-1,k+1,\cdots n} = \frac{-R_{jk}}{\sqrt{R_{jj} R_{kk}}} .$$

University of Minnesota.

NOTE ON REGRESSION FUNCTIONS IN THE CASE OF THREE SECOND ORDER RANDOM VARIABLES

By Clyde A. Bridger

The study of the correlation of two second-order random variables has received the attention of several authors, among them Yule [1], Charlier [2], Wicksell [3, 4], and Tschuprow [5]. Yule writes of them under the guise of "attributes." The study of three or more second order random variables has lagged behind. In this note we shall examine the regression function of one second order random variable on two others by considering the problem from the point of view of Tschuprow's [6] paper on the correlation of three random variables.

A variable X that takes on m values $x_1, \cdots, x_m$ with corresponding probabilities $p_1, \cdots, p_m$ subject to the condition $\sum_i p_i = 1$ is defined as a random variable of order m. (In particular, if X takes on only two values, x and x' with probabilities p and q, where $p + q = 1$, X is a random variable of second order.) The system of values x and probabilities p constitute the law of distribution of X. In the case of two random variables, X and Y, there exists a joint distribution law, covering all possible combinations of X and Y, together with their associated probabilities $p_{11}, \cdots, p_{mn}$ the joint distribution law contains all of the information regarding the stochastical dependence of X and Y.

The extension to more than two variables is immediate. Let p_{ijk} represent the probability of the simultaneous occurrence of the set of values x_i, y_j, z_k of three random variables X, Y, and Z; p_{ij}, that of the simultaneous occurrence of x_i, y_j together without reference to Z; p_i, that of the occurrence of x_i without reference to Y or Z; etc. Then, we have relationships of the types $\sum_i \sum_j \sum_k p_{ijk} = \sum_i \sum_j p_{ij} = \sum_i p_i = 1$; $\sum_i p_{ijk} = p_{jk}$; $\sum_i \sum_j p_{ijk} = \sum_j p_{jk} = \sum_i p_{ik} = p_k$. Similarly, let $p_{jk}^{(i)}$ be the probability of the simultaneous occurrence of y_j and z_k on the condition that X takes on the value x_i; $p_j^{(i)}$, that of the occurrence of y_j without reference to Z, on the same condition; etc. Then

$$\sum_k p_k^{(i)} = \sum_k p_k^{(ij)} = \sum_j \sum_k p_{jk}^{(i)} = 1; \qquad \sum_j p_{jk}^{(i)} = p_k^{(i)}; \qquad p_i p_j^{(i)} = p_{ij};$$

$$p_{ij} p_k^{(j)} = p_i p_{jk}^{(i)} = p_i p_j^{(i)} p_k^{(ij)} = p_{ijk}; \qquad \sum_i p_i p_j^{(i)} = p_j; \text{ etc.}$$

Denoting by $E(x)$ or simply Ex the expression "the mean value or mathematical expectation of x," we have $m_{fgh} = EX^f Y^g Z^h = \sum_i \sum_j \sum_k p_{ijk} x_i^f y_j^g z_k^h$. In particular, the mean values of the distributions are given by $m_X = EX$

$= \sum_i p_i x_i$, $m_Y = EY = \sum_j p_j y_j$, $m_Z = EZ = \sum_k p_k y_k$. Then we may write $\mu_{fgh} = E(X - m_X)^f(Y - m_Y)^g(Z - m_Z)^h = Eu^f v^g w^h = \sum_i \sum_j \sum_k p_{ijk}(x_i - m_X)^f (y_j - m_Y)^g(z_k - m_Z)^h$. The quantities μ may be identified as terms in the expression for the moments for the sum of three variables as follows: $E(u + v + w)^n = Eu^n + nEu^{n-1}v + \cdots + kEu^f v^g w^h + \cdots + Ew^n$, where $f + g + h = n$. If $n = 2$, we have the variance of the sum of three variables given by $\mu_{2..} + 2\mu_{11.} + \mu_{.2.} + 2\mu_{.11} + 2\mu_{1.1} + \mu_{..2}$, where the dots in the subscripts indicate variables not considered. Thus $\mu_{2..}$ refers to the second moment of the distribution of the variable X about its mean, m_X, without consideration of the distributions of Y or Z. If every term of the expansion of the n-th moment of the sum of three variables is divided by the quantity $\sqrt{\mu_{2..}{}^f\mu_{.2.}{}^g\mu_{..2}{}^h}$, the expansion takes the "normal form." The type term is $r_{fgh} = \mu_{fgh}/\sqrt{\mu_{2..}{}^f\mu_{.2.}{}^g\mu_{..2}{}^h}$. In the case of one variable, $r_f = \mu_f/\sqrt{\mu_2{}^f}$, so $r_1 = 0$, $r_2 = 1$, $r_3 = \sqrt{\beta_1}$, $r_4 = \beta_2$, etc. In the case of two variables, $r_{1.} = r_{.1} = 0$, $r_{2.} = r_{.2} = 1$, r_{11} = Pearson's product-moment coefficient of correlation, etc. Functions of parameters r will serve to characterize the law of correlation among the variables.

By writing the expressions with superscript (i) to denote that the values of the distributions of Y and Z are those which correspond to a fixed value x_i of the distribution of X, we have $m_Y{}^{(i)} = (EY)^{(i)}$, $m_Z{}^{(i)} = (EZ)^{(i)}$, $\mu_{gh}{}^{(i)} = E(Y - m_Y{}^{(i)})$ times $(Z - m_Z{}^{(i)})$, $r_{gh}{}^{(i)} = \mu_{gh}{}^{(i)}/\sqrt{\mu_{.2.}^{(i)g}\,\mu_{..2}^{(i)h}}$. (For $g = h = 1$, $r_{gh}{}^{(i)}$ becomes the conditional coefficient of correlation between Y and Z for $X = x_i$.) Thus it follows that we can study the correlation between Y and Z for each value of X separately.

For second order random variables, some changes in notation can be made. Let p_x and $p_{x'}$ be the probabilities corresponding to the values x and x', respectively, of X; p_y and $p_{y'}$ correspond to y and y', respectively; p_z and $p_{z'}$ correspond to z and z', respectively. Also, let p_{xy} represent the probability of the simultaneous occurrence of x and y together without reference to the distribution of Z, etc., and p_{xyz} represent the probability of the simultaneous occurrence of all three values, x, y, z, of their respective distributions, etc. Then, $p_x + p_{x'} = p_y + p_{y'} = p_z + p_{z'} = 1$; $p_{xy} + p_{xy'} = p_x$; $p_{xyz} + p_{xyz'} + p_{xy'z} + p_{xy'z'} = p_x$; etc.

Let us set up a system of normal coordinates in which the values U_i along the U-axis are defined by $U_i = \dfrac{x_i - m_X}{\sqrt{\mu_{2..}}}$, those along the V-axis by $V_j = \dfrac{y_j - m_Y}{\sqrt{\mu_{.2.}}}$, and those along the W-axis by $W_k = \dfrac{z_k - m_Z}{\sqrt{\mu_{..2}}}$. Let $m_Z{}^{(ij)}$ represent the mean of the set of values of the Z distribution which correspond to the fixed pair of values, (x_i, y_j), of the X and Y distributions. Then, in the new coordinate system, the same thing is given by $M_W{}^{(ij)} = \dfrac{m_Z{}^{(ij)} - m_Z}{\sqrt{\mu_{..2}}}$. Now, the series of values $M_W{}^{(ij)}$ obtained by giving i and j different values for the pair (U_i, V_j) determine what is called the regression function of W on U and V (or, in the

original notation, the surface of regression of the distribution of Z on the distributions of X and Y). Similarly, the values of $[M_W^{(ij)}]^{(i)} = \frac{m_Z^{(ij)} - m_Z^{(i)}}{\sqrt{\mu_{\cdot\cdot 2}}}$ obtained by fixing U and varying V in the set (U_i, V_j) determine what is called the conditional line of regression of W on V for a fixed value of U. With these definitions we shall consider the problem of finding a regression function of W on U and V for three second order random variables.

For convenience, write $\delta_{xy} = p_{xy} - p_x p_y$, $\delta_{xz} = p_{xz} - p_x p_z$, $\delta_{yz} = p_{yz} - p_y p_z$, $\alpha_{yz} = p_x p_{xyz} - p_{xy} - p_{xz}$, $\epsilon_z = p_{xyz} - p_{xy} p_z$, $\beta_{yz} = p_{x'} p_{x'yz} - p_{x'y} p_{x'z}$, $\theta_{xyz} = \epsilon_z - p_y \delta_{xz} - p_x \delta_{yz} = \epsilon_y - p_x \delta_{yz} - p_z \delta_{xy} = \epsilon_x - p_y \delta_{xz} - p_z \delta_{xy}$. Direct substitutions into the several formulas developed above then gives us the representative forms to be used in subsequent calculations:

$$x - m_X = p_{x'}(x - x'), \quad x' - m_X = -p_x(x - x').$$

$$m_X = p_x x + p_{x'} x', \quad r_{1\cdot\cdot} = 0, \quad r_{2\cdot\cdot} = 1, \quad r_{3\cdot\cdot} = \frac{p_{x'} - p_x}{\sqrt{p_x p_{x'}}},$$

$$r_{4\cdot\cdot} = \frac{1}{p_x p_{x'}} - 3, \quad r_{11\cdot} = \frac{\delta_{xy}}{\sqrt{p_x p_{x'} p_y p_{y'}}}, \quad r_{21\cdot} = r_{11\cdot} r_{3\cdot\cdot},$$

$$r_{12\cdot} = r_{11\cdot} r_{\cdot 3\cdot}, \quad r_{13\cdot} = r_{11\cdot} r_{\cdot 4\cdot}, \quad r_{22\cdot} = r_{11\cdot} r_{3\cdot\cdot} r_{\cdot 3\cdot} + 1,$$

$$r_{12\cdot} r_{21\cdot} = r_{11\cdot}(r_{22\cdot} - 1), \quad r_{211} = r_{3\cdot\cdot} r_{111} + r_{\cdot 11},$$

$$r_{121} = r_{\cdot 3\cdot} r_{111} + r_{1\cdot 1}, \quad r_{112} = r_{\cdot\cdot 3} r_{111} + r_{11\cdot},$$

$$r_{111} = \frac{\theta_{xyz}}{\sqrt{p_x p_{x'} p_y p_{y'} p_z p_{z'}}}, \quad U_1 = \frac{p_{x'}}{\sqrt{p_x p_{x'}}}, \quad U_2 = \frac{-p_x}{\sqrt{p_x p_{x'}}},$$

$$M_W^{(11\cdot)} = \frac{\epsilon_z}{p_{xy}\sqrt{p_z p_{z'}}}, \quad M_W^{(12\cdot)} = \frac{\delta_{xz} - \epsilon_z}{p_{xy'}\sqrt{p_z p_{z'}}},$$

$$M_W^{(21\cdot)} = \frac{\delta_{yz} - \epsilon_z}{p_{x'y}\sqrt{p_z p_{z'}}}, \quad M_W^{(22\cdot)} = \frac{\epsilon_z - \delta_{xz} - \delta_{yz}}{p_{x'y'}\sqrt{p_z p_{z'}}},$$

$$[M_W^{(11\cdot)}]^{(1\cdot\cdot)} = \frac{\alpha_{yz}}{p_{xy}\sqrt{p_{xz} p_{xz'}}}, \quad [M_W^{(21\cdot)}]^{(2\cdot\cdot)} = \frac{\beta_{yz}}{p_{x'y}\sqrt{p_{x'z} p_{x'z'}}},$$

$$[M_W^{(12\cdot)}]^{(1\cdot\cdot)} = \frac{-\alpha_{yz}}{p_{xy'}\sqrt{p_{xz} p_{xz'}}}, \quad [M_W^{(22\cdot)}]^{(2\cdot\cdot)} = \frac{-\beta_{yz}}{p_{x'y'}\sqrt{p_{x'z} p_{x'z'}}}.$$

In the case of correlation of two second order random variables, a linear regression function can always be found [3, 5]. Similarly, the conditional regression functions in the case of three second order random variables can always be taken as linear. If we take as the form of the regression function of W on U and V the form $M_W^{(ij)} = aU_i + bV_j + cU_iV_j + d$, where a, b, c, d are constants to be determined by direct substitution for U_i and V_j from the distributions of X and Y, it is seen that linearity of all total and conditional

regression functions is preserved. By total regression function, we mean the regression of W on U or W on V.

Now consider the problem of finding a, b, c, d. The direct substitution provides us with four linearly dependent equations in four unknowns. Linear combinations reduce the set to three, from which the relationship $d = -cr_{11\cdot}$ is obtained. By building up the various terms in the equations through dividing by the necessary values of p, the parameters r can be made to appear. Further combinations now reduce the set to the following three:

$$\begin{aligned} r_{111} &= ar_{21\cdot} + br_{12\cdot} + c(r_{22\cdot} - r_{11\cdot}^2) \\ r_{\cdot 11} &= ar_{11\cdot} + b \quad\; + cr_{12\cdot} \\ r_{1\cdot 1} &= a \quad\;\; + br_{11\cdot} + cr_{21\cdot} \end{aligned}$$

The solution gives

$$a = \frac{r_{1\cdot 1} - r_{11\cdot}r_{\cdot 11}}{1 - r_{11\cdot}^2} - \frac{r_{21\cdot} - r_{11\cdot}r_{12\cdot}}{1 - r_{11\cdot}^2}c = a' - a''c$$

$$b = \frac{r_{\cdot 11} - r_{11\cdot}r_{1\cdot 1}}{1 - r_{11\cdot}^2} - \frac{r_{12\cdot} - r_{11\cdot}r_{21\cdot}}{1 - r_{11\cdot}^2}c = b' - b''c$$

$$c = (1 - r_{11\cdot}^2)(r_{111} - a'r_{12\cdot} - b'r_{21\cdot}) \div \Delta, \qquad \text{where}$$

$$\Delta = \begin{vmatrix} 1 & r_{11\cdot} & r_{21\cdot} \\ r_{11\cdot} & 1 & r_{12\cdot} \\ r_{21\cdot} & r_{12\cdot} & r_{22\cdot} - r_{11\cdot}^2 \end{vmatrix}$$

The regression function becomes
$M_W^{(ij)} = a'U_i + b'V_j - c(r_{111} + a''U_i + b''V_j - U_iV_j)$. If $c = 0$ the surface is a plane. Examination of the characteristics of r_{111} shows that generally c cannot be zero. The vanishing of c implies that special relations must exist between p_{ijk} and p_{ij}, p_{ik}, p_{jk}.

Two constants of considerable importance in the theory of correlation are the multiple correlation coefficient and the multiple correlation ratio. For the regression of W on U and V, the former is defined as $R_{11\cdot}^2 = a'r_{1\cdot 1} + b'r_{\cdot 11}$ and the latter as $\eta_{--2} = \sum_i \sum_j p_{ij}[M_W^{(ij)}]^2$. For planar regression, the difference $\eta_{--2} - R_{11\cdot}^2$ must vanish. For others, the difference takes on values characteristic of the regression function. To find the value it takes for our case, we set up the value of η_{--2} from the regression function just given and subtract $R_{11\cdot}^2$.

By direct substitution, we have $\eta_{--2} - R_{11\cdot}^2 = \sum_i \sum_j p_{ij}(aU_i + bV_j - cU_iV_j - cr_{111})^2 - a'r_{1\cdot 1} - b'r_{\cdot 11}$. Since $\sum_i \sum_j p_{ij}U_i^2 = 1$, $\sum_i \sum_j p_{ij}(U_iV_j)^2 = r_{22\cdot}$, etc., we find rather easily that

$$\eta_{--2} - R_{11\cdot}^2 = c^2(r_{22\cdot} - r_{11\cdot}^2) - a''r_{21\cdot} - b''r_{12\cdot}.$$

We can also obtain the same value of $\eta_{--2} - R_{11\cdot}^{2}$ by direct substitution for the four values of $M_W^{(ij)}$ in η_{--2} and subtracting $R_{11\cdot}^{2}$. To actually obtain this is a long laborious process complicated by the fact that so many alternate forms for the answer are possible, of which only one is comparable with the value previously found. The general procedure is first to set up from the definition the expression $K = p_z p_{z'}\, \eta_{--2} =$

$$p_{xy}\left(\frac{\epsilon_z}{p_{xy}}\right)^2 + p_{xy'}\left(\frac{\delta_{xz} - \epsilon_z}{p_{xy'}}\right)^2 + p_{x'y}\left(\frac{\delta_{yz} - \epsilon_z}{p_{x'y}}\right)^2 + p_{x'y'}\left(\frac{\epsilon_z - \delta_{xz} - \delta_{yz}}{p_{x'y'}}\right)^2.$$

Then we build up each square by addition and subtraction so that it will contain a θ_{xyz} term. At the close of the process, we convert the whole expression into the parameters r by dividing through by $p_z p_{z'}(p_x p_{x'} p_y p_{y'})^2$ and substituting from the list of representative forms given at the beginning of the paper. A matter of rearrangement now gives the same result as before.

From the symmetry involved, we can say that, in the case of the correlation of three second order random variables, the function representing the regression of one on the other two has an equation in normal coordinates of the form $M_W^{(ij)} = aU_i + bV_j - cU_iV_j - cr_{111}$, where a, b, and c satisfy equations of type

$$\begin{aligned} r_{111} &= ar_{21\cdot} + br_{12\cdot} + c(r_{22\cdot} - r_{11\cdot}^{2}) \\ r_{\cdot 11} &= ar_{11\cdot} + b \qquad + cr_{12\cdot} \\ r_{1\cdot 1} &= a \qquad + br_{11\cdot} + cr_{21\cdot} \end{aligned}$$

State Division of Public Health,
Boise, Idaho.

BIBLIOGRAPHY

[1] G. Udny Yule. *An Introduction to the Theory of Statistics.* London: Charles Griffin & Co., Ltd., 1922. 6th Ed.

[2] C. V. L. Charlier. "Om korrelation mellan eganskaper inom den homograda statistiken." *Svenska Aktuarieföreningens Tidskrift.* Vol. I (1914), pp. 21–35.

[3] S. D. Wicksell. "Some theorems in the theory of probability, with special reference to their importance in the theory of homograde correlation." *Svenska Aktuarieföreningens Tidskrift.* Vol. III (1916), pp. 165–213.

[4] S. D. Wicksell. "On the correlation of acting probabilities." *Skandinavisk Aktuarietidskrift.* Vol. I (1918), pp. 98–135.

[5] A. A. Tschuprow. *Grundbegriffe und Grundprobleme der Korrelationstheorie.* Leipzig: B. G. Teubner, 1925.

[6] A. A. Tschuprow. (Translation into English by L. Isserlis.) "The Mathematical Theory of the Statistical Methods Employed in the Study of Correlation in the Case of Three Variables." *Transactions of the Cambridge Philosophical Society.* Vol. XXIII, no. 12 (1928), pp. 337–382.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

SEPTEMBER, 1938 · $1.50 PER COPY · $6.00 PER ANNUM · VOL. 33 · NO. 203

CONTENTS

ARTICLES

Address inquiries and orders for subscriptions and back numbers to Frederick F. Stephan, Secretary, American Statistical Association, 722 Woodward Building, Washington, D. C.

THE INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute of Mathematical Statistics is to stimulate research in the mathematical theory of statistics and to promote coöperation between the field of pure research and the fields of application.

Membership dues including subscription to the Annals of Mathematical Statistics are $5.00 per year. The dues and all correspondence regarding membership in the Institute should be sent to the Secretary-Treasurer of the Institute.